BEST AVAILABLE COPY

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-067187
(43)Date of publication of application : 16.03.2001

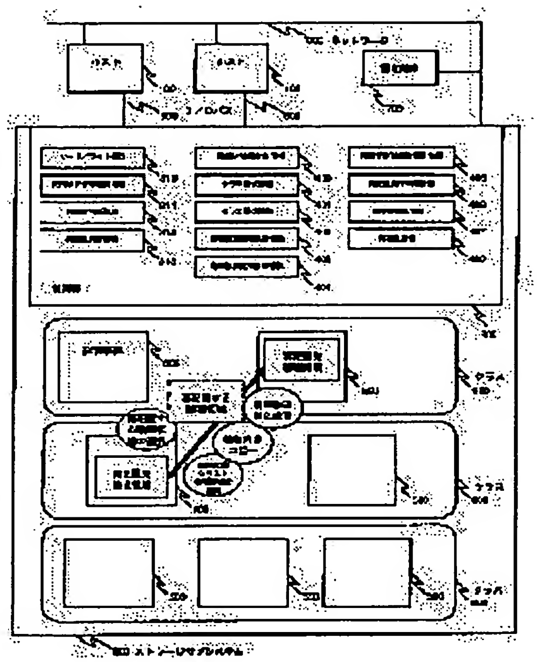
(51)Int.Cl. G06F 3/06
G06F 12/00

(21)Application number : 11-242713
(22)Date of filing : 30.08.1999

(71)Applicant : HITACHI LTD
(72)Inventor : ARAKAWA TAKASHI
MOGI KAZUHIKO
YAMAKAMI KENJI
ARAI HIROHARU

(54) STORAGE SUB-SYSTEM AND ITS CONTROL METHOD

(57)Abstract:



PROBLEM TO BE SOLVED: To simplify a work for optimizing arrangement by re-arrangement by the user of a disk array system or the like by changing the correspondence of a logical storage area from a physical storage area into the second physical storage area and executing re-arrangement.

SOLUTION: A control part 300 automatically executes re-arrangement execution processing at the set time and date. That is, the part 300 copies contents stored in a re-arrangement source physical area in a re-arrangement destination physical area based on re-arrangement information 408. Moreover, at the point of time when the copying is completed and the whole contents of the re-arrangement source physical area are reflected in the re-arrangement destination physical area, the control part 300 changes a physical area corresponding to a logical area for executing re-arrangement in logical/physical correspondence information 400 from the re-arrangement source physical area into the re-arrangement destination physical area. Besides, the control part 300 uses the re-arrangement destination physical area on a non-usage physical area 1470, changes the re-arrangement source physical area into the non-usage one and, moreover, updates the time and date of re-arrangement execution time information 406 into the one for a next time by referring to time and date updating information on re-arrangement execution time information 406.

LEGAL STATUS

[Date of request for examination] 18.06.2002

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]	
[Patent number]	3541744
[Date of registration]	09.04.2004
[Number of appeal against examiner's decision of rejection]	
[Date of requesting appeal against examiner's decision of rejection]	
[Date of extinction of right]	

Copyright (C); 1998,2003 Japan Patent Office

* NOTICES *

JPO and NCIPI are not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]
[Claim 1] Two or more storage and a means to acquire the operating condition information on said storage, It has a means to perform matching with the logic storage region which said computer makes a read/write object, and the first physical memory field of said storage. It is the control approach of the storage subsystem linked to one or more calculating machines. Said storage is classified into two or more groups (class), and said class has the set-up attribute. Said storage subsystem Based on said operating condition information and said class attribute, the class of the suitable relocation place for said logic storage region is determined. The second physical memory field available as a relocation place of said logic storage region is chosen from the inside of said class. The control approach of said storage subsystem characterized by rearranging by changing matching of a logic storage region into said second physical memory field from said first physical memory field while copying the contents of said first physical memory field to said said second physical memory field.
[Claim 2] It is the control approach of the storage subsystem which it is the control approach of a storage subsystem according to claim 1, and a storage subsystem accumulates said operating condition information, and determines the relocation place of a logic storage region and is characterized by rearranging to the set-up time amount based on said operating condition information on the set-up period.
[Claim 3] It is the control approach of a storage subsystem according to claim 1 or 2. A storage subsystem As operating condition information, the time per unit time amount of storage (activity ratio) is used. Each class It has the engine-performance ranking and the activity ratio upper limit between the classes set up as an attribute. Said storage subsystem The control approach of the storage subsystem characterized by choosing the logic storage region rearranged from the storage exceeding the activity ratio upper limit of a class, and determining that the class of the relocation place of said logic storage region will not exceed the activity ratio upper limit of a class to each class of the high order of said ranking.
[Claim 4] It is the control approach of a storage subsystem according to claim 1 or 2. A storage subsystem As operating condition information, the time per unit time amount of storage (activity ratio) is used. Each class It has the engine-performance ranking and the activity ratio upper limit between the classes set up as an attribute. Said storage subsystem The logic storage region rearranged from the storage exceeding the activity ratio upper limit of a class is chosen. The control approach of the storage subsystem characterized by determining that a physical memory field available as a relocation place of said logic storage region will not exceed the activity ratio upper limit of said class from the storage in the same class.
[Claim 5] It is the control approach of a storage subsystem according to claim 1 or 2. A storage subsystem As operating condition information, the time per unit time amount of storage (activity ratio) is used. Each class has the object access classification and the activity ratio upper limit which were set up as an attribute. Said storage subsystem The logic storage region rearranged from the storage exceeding the activity ratio upper limit of a class is chosen. The control approach of the storage subsystem characterized by determining that the class of the relocation place of said logic storage region will not exceed the activity ratio upper limit of a class to each

class of said object access classification based on the analysis result of the access classification to said logic storage region.
[Claim 6] A means to connect with one or more computers and to acquire the operating condition information on two or more storage and said storage, It is the storage subsystem which has a means to perform matching with the logic storage region which said computer makes a read/write object, and the first physical memory field of said store. A means to manage said two or more disk units as two or more groups (class) which have an attribute, respectively, A means to determine the class of the suitable relocation place for said logic storage region based on said operating condition information and said class attribute, A means to choose the second physical memory field available as a relocation place of said logic storage region from the inside of said class, The storage subsystem characterized by having the means which rearranges by changing matching of a logic storage region into said second physical memory field from said first physical memory field while copying the contents of said first physical memory field to said said second physical memory field.
[Claim 7] It is the storage subsystem which is a storage subsystem according to claim 6, and is characterized by having a means for a storage subsystem to accumulate said operating condition information, and to determine the relocation place of a logic storage region automatically based on said operating condition information on the set-up period, and the means which rearranges to the set-up time amount.
[Claim 8] It is a storage subsystem according to claim 6 or 7. A storage subsystem It has a means using the time per unit time amount of a store (activity ratio) as operating condition information. Said storage subsystem A means to choose the logic storage region rearranged from the storage exceeding the activity ratio upper limit set as each class as an attribute. The storage subsystem characterized by having a means to determine not to exceed the activity ratio upper limit of each class from the engine-performance ranking between the classes set as each class as an attribute in the class of the relocation place of said logic storage region.
[Claim 9] It is a storage subsystem according to claim 6 or 7. A storage subsystem It has a means using the time per unit time amount of a store (activity ratio) as operating condition information. Said storage subsystem A means to choose the logic storage region rearranged from the storage exceeding the activity ratio upper limit of the class set up as an attribute, A means to analyze the access classification to said logic storage region, and object access classification, from the class set up as an attribute The storage subsystem characterized by having a means to determine that the class of the relocation place of said logic storage region will not exceed the activity ratio upper limit of each class based on said analysis result.
[Claim 10] It is the storage subsystem characterized by being a storage subsystem given in claims 6, 7, 8, or 9, and for a storage subsystem being a disk array which has two or more disk units, and having a means using the activity ratio of said disk unit as operating condition information.

[Translation done.]

* NOTICES *

JPO and NCIPI are not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention] This invention relates to the storage subsystem which has two or more stores, and its control approach.
[0002]
[Description of the Prior Art] In a computer system, a disk array system is in one of the secondary-storage systems which realizes high performance. A disk array system is a system which performs read/write of the data by which arrange two or more disk units in the shape of an array, and division storing is carried out at said each disk unit at a high speed by operating said each disk unit to juxtaposition. As a paper about a disk array system, it is D.A.Patterson. G.Gibson and There are R.H.Kats and "A Case for Redundant Arrays of Inexpensive Disks (RAID)" (in Proc. ACM SIGMOD, pp.109-116, June 1988). In this paper, the classification of level 5 is given from level 1 to the disk array system which added redundancy according to that configuration. In addition to these classification, a disk array system without redundancy may be called level 0. In building a disk array system, since cost, performance characteristics, etc. for redundancy etc. to realize differ from each other, each above-mentioned level makes the array (group of a disk unit) of two or more level intermingled in many cases. Here, this group is called a parity group.
[0003] In order to realize optimal cost performance in cost's changing with the engine performance, capacity, etc. and building a disk array system, two or more sorts of disk units from which the engine performance and capacity differ too may be used for a disk unit.
[0004] In order to distribute and arrange the data stored in a disk array system to a disk unit as mentioned above, a disk array system matches the physical memory field which shows the logic storage region which the host computer linked to a disk array system accesses, and the storage region of a disk unit (address translation). The disk array system which realizes the optimal arrangement of the stored data is indicated by JP,9-274544,A with a means to acquire the information about I/O access over the logic storage region from a host computer, and a means to change matching with the physical memory field of a logic storage region, and to perform physical relocation.
[0005]
[Problem(s) to be Solved by the Invention] The following technical problems occur about the activation approach of the arrangement optimization in a Prior art as shown in JP,9-274544,A.
[0006] In selection of the logic storage region to rearrange, and selection of the physical memory field of a relocation place, the user or customer engineer of a disk array system had to check information, such as said disk array.structure of a system, property of each disk unit, engine performance, etc., and had to perform said selection, and the activity by the user or the customer engineer was complicated.
[0007] Moreover, when a disk array system chose automatically, the user or the customer engineer had to check the information on said each disk unit, the selection-criterion value had to be specified, and the activity by the user or the customer engineer was complicated too. The complicatedness of information management increases especially about the disk array system by

which level of a different kind and a disk unit of a different kind are intermingled as mentioned above.
[0008] Moreover, reference of the I/O access information of selection of a disk array system performed for accumulating was not taking into consideration the property of the schedule of processing performed by the system containing a host computer and a disk array system. I/O accompanying the processing and processing which are generally performed by the computer system is performed in conformity with the schedule created by the user, and the inclination of processing and I/O shows the periodicity for every month and every year day by day in many cases, and, generally a user is considered to be interested in processing and I/O of a specific period.
[0009] Moreover, in the above-mentioned conventional technique, the following technical problems occur about the engine-performance tuning approach by relocation. Although the engine-performance tuning approach by physical relocation adds modification to the operating condition of a disk unit, i.e., a physical memory field, since it referred to the information about I/O access over the logic storage region from a host computer, it may be unable to perform right selection in the Prior art in selection of the logic storage region to rearrange, and selection of the physical memory field of a relocation place.
[0010] Moreover, even when the sequential access and random access from a host computer are performed to the separate physical memory field notably included in the same disk unit, in order to divide a sequential access and random access into a different disk unit, the disk unit of a relocation place was able to be specified as arbitration, and automatic relocation was not able to be made to perform. Generally, although random access with a small data length is asked for a response (high response engine performance) in a short time as requirements for processing from a host computer, when a sequential access with a large data length exists in the same disk unit, the response time of random access will be checked by processing of a sequential access, and will become long, and the response engine performance will get worse.
[0011] The first purpose of this invention is to do simple an activity for the user or customer engineer of a disk array system to perform arrangement optimization by relocation.
[0012] The second purpose of this invention is to enable arrangement optimization by relocation in consideration of the schedule of processing by the system containing a host computer and a disk array system.
[0013] The third purpose of this invention is to offer the control approach of a disk array system and disk array system which perform selection based on the operating condition of the disk unit which is an actual store in selection of the logic storage region to rearrange, and selection of the physical memory field of a relocation place.
[0014] The fourth purpose of this invention is to enable it to separate into the disk unit which specifies the disk unit of a relocation place as arbitration, and changes a sequential access and random access with relocation automatically to the mixture of the remarkable sequential access in the same disk unit, and random access in a disk array system.
[0015]
[Means for Solving the Problem] In order to realize the first above-mentioned purpose, the disk array system linked to one or more sets of host computers It has a means to perform matching with a means to acquire the operating condition information on two or more disk units of a subordinate, and the logic storage region and the first physical memory field of a disk unit which a host computer makes a read/write object. Furthermore, a means to manage two or more disk units as two or more groups (class) which have an attribute, respectively, A means to determine the class of the suitable relocation place for a logic storage region based on operating condition information and a class attribute, A means to choose the second physical memory field available as a relocation place of a logic storage region from the inside of a class, While copying the contents of the first physical memory field to said said second physical memory field, it has the means which rearranges by changing matching of a logic storage region into the second physical memory field from the first physical memory field.
[0016] Moreover, in order to realize the second purpose of the above, a disk array system can be equipped with a means to accumulate operating condition information and to determine the

relocation place of a logic storage region based on the operating condition information on the set-up period, and the means which rearranges to the set-up time amount.
[0017] Moreover, in order to realize the third purpose of the above, a disk array system is equipped with a means to use the time per unit time amount of a disk unit (activity ratio), as operating condition information.
[0018] In order to realize the fourth purpose of the above, moreover, a disk array system The object access classification (sequential / random access classification) and the activity ratio upper limit which were set as each class as an attribute are used. The logic storage region rearranged from the storage exceeding the activity ratio upper limit of a class is chosen. Based on the analysis result of the access classification to a logic storage region, it has a means to determine that the class of the relocation place of a logic storage region will not exceed the activity ratio upper limit of a class to each class of a suitable access classification.
[0019]
[Embodiment of the Invention] Hereafter, the gestalt of operation of this invention is explained using drawing 1 - drawing 27 .
[0020] The gestalt of <gestalt of the first operation> book operation explains the scheduling of decision of the relocation based on a class 600, relocation decision, and activation.
[0021] Drawing 1 is the block diagram of the computing system in the gestalt of operation of the 1st of this invention.
[0022] The computer system in the gestalt of this operation comes to have a host 100, the storage subsystem 200, and a control terminal 700.
[0023] A host 100 connects with the storage subsystem 200 through I/O bus 800, and performs I/O of a lead or a light to the storage subsystem 200. A host 100 specifies a logic field about the storage region of the storage subsystem 200 in the case of I/O. There are ESCON, SCSI, a fiber channel, etc. as an example of I/O bus 800.
[0024] The storage subsystem 200 has a control section 300 and two or more storage 500. A control section 300 performs the read/write processing 310, the operating condition information acquisition processing 311, the relocation decision processing 312, and relocation executive operation 313. Moreover, the storage subsystem 200 holds the information 400 corresponding to logic/physics, the class configuration information 401, the class attribute information 402, the logic field operating condition information 403, the physical field operating condition information 404, the relocation decision horizon information 405, the relocation activation time information 406, the free-space information 407, and relocation information 408.
[0025] A host 100, a control section 300, and a control terminal 700 are connected in a network 900. There are FDDI, a fiber channel, etc. as an example of a network 900.
[0026] Although components generally used in a computer, such as memory for performing processing in each and CPU, also exist in a host 100, a control section 300, and a control terminal 700, respectively, since it is not important, explanation is omitted in explanation of the gestalt of this operation here.
[0027] A host 100 explains the read/write processing 310 in the case of performing read/write to the storage subsystem 200, and the operating condition information acquisition processing 311 by drawing 2 .
[0028] In the read/write processing 310, from the control section 300 of the storage subsystem 200, a host 100 specifies a logic field and demands a lead or a light (step 1000). The control section 300 which received the demand asks for the physical field corresponding to a logic field using the information 400 corresponding to logic/physics, namely, changes the address (logical address) of a logic field into the address (physical address) of a physical field (step 1010). Then, in a lead, data are read from the store 500 of this physical address, and a control section 300 transmits it to a host 100, in the case of a light, stores in the store 500 of said physical address the data transmitted by the host 100 (step 1020), and performs the further below-mentioned operating condition information acquisition processing 311. Read/write demand and data transfer are performed through I/O bus 800.
[0029] An example of the information 400 corresponding to logic/physics is shown in drawing 3 . The logical address is the address which shows the logic field which a host 100 uses by the

read/write processing 310. A physical address is the address which shows the field on the storage 500 with which data are actually stored, and consists of a storage number and the address in storage. A storage number shows each storage 500. The address in storage is the address which shows the storage region within storage 500.
[0030] Next, in the operating condition information acquisition processing 311, a control section 300 updates the logic field operating condition information 403 about the logic field which became a read/write object in the read/write processing 310, and the physical field operating condition information 404 about the physical field used by the read/write processing 310 (steps 1030 and 1040). The logic field operating condition information 403 and the physical field operating condition information 404 are the information about operating conditions of each time of each logic field and physical field, such as for example, operating frequency, an activity ratio, and an attribute about read/write. The gestalt of subsequent operations explains the concrete example of the logic field operating condition information 403 and the physical field operating condition information 404.
[0031] Next, drawing 4 explains the relocation decision processing 312 which a control section 300 performs.
[0032] Storage 500 is classified into two or more groups (class 600) as a user or an initial state, and the classification to a class 600 is set as the class configuration information 401. Furthermore, each class 600 is having the attribute set up as a user or initial condition, and the attribute is set as the class attribute information 402. The class attribute information 402 is the information about attributes, such as a permissible operating condition, a suitable operating condition, and priority between classes. The gestalt of subsequent operations explains the concrete example of the class configuration information 401 and the class attribute information 402. the relocation decision horizon information 405 -- a user -- or the period and period update information of operating condition information which are made into the object of the relocation decision processing 312 as initial condition are set up.
[0033] An example of the relocation decision horizon information 405 is shown in drawing 5 R> 5. The period from initiation time to termination time turns into a horizon. Period update information is the setups of a next horizon, for example, may have X time amount back etc. every week and every day. A control section 300 chooses the logic field which should perform physical relocation as compared with the permissible operating condition of each class 600 of the class attribute information 402 (step 1110) etc. with reference to the logic field operating condition information 403 on a horizon, and the physical field operating condition information 404 (step 1100) (step 1120).
[0034] Furthermore, with reference to the permissible operating condition and the suitable operating condition of the class attribute information 402, the priority between classes (step 1130), etc., a control section 300 chooses the class 600 of the relocation place of a logic field (step 1140), chooses a physical field intact as a relocation place of a logic field from the storage 500 belonging to a class 600 further (step 1150), and outputs a selection result to relocation information 408 (step 1160).
[0035] An example of relocation information 408 is shown in drawing 6 . A logic field is a logic field to rearrange, rearranging agency physics fields are the storage number which shows the current physical field corresponding to a logic field, and the address in storage, and relocation place physics fields are the storage number which shows the physical field of a relocation place, and the address in storage. As shown in drawing 6 , one or more plannings of relocation are performed and it gets. Furthermore, a control section 300 updates the horizon of the relocation decision horizon information 405 to degree batch with reference to the period update information of the relocation decision horizon information 405 (step 1170). In the above-mentioned processing, a control section 300 uses the free-space information 407 for retrieval of the aforementioned intact physical field, using the information 400 corresponding to logic/physics.
[0036] An example of the free-space information 407 is shown in drawing 7 . A storage number shows each storage 500. The address in storage is the address which shows the field within storage 500. A storage number and the address in equipment show a physical field, and use / intact item shows use / intact distinction of a physical field. A control section 300 usually

performs relocation decision processing 312 automatically before the below-mentioned relocation executive operation 313 after a horizon.
[0037] Next, drawing 8 explains the relocation executive operation 313 which a control section 300 performs.
[0038] the relocation activation time information 406 — a user -- or the time and time update information which perform relocation executive operation 313 as initial condition are set up.
[0039] An example of the relocation activation time information 406 is shown in drawing 9 . A control section 300 performs automatically relocation executive operation 313 explained below in the set-up time. Time update information is setups of time which perform next relocation executive operation 313, for example, may have X time amount back etc. every week and every day. A control section 300 copies the contents stored in a rearranging agency physics field based on relocation information 408 to a relocation place physics field (step 1200). Furthermore, when a copy is completed and all the contents of the rearranging agency physics field are reflected in a relocation place physics field, a control section 300 changes into a relocation place physics field the physical field corresponding to the logic field which performs relocation on the information 400 corresponding to logic/physics from a rearranging agency physics field (step 1210).
[0040] Furthermore, a control section 300 considers the relocation place physics field on the intact physics field 470 as use, and changes a rearranging agency physics field intact (step 1220). Furthermore, a control section 300 updates the time of the relocation activation time information 406 to degree batch with reference to the time update information of the relocation activation time information 406 (step 1230).
[0041] A user or a customer engineer can check and set up minding a network 900 from a control terminal 700, or setting up and checking each information which the control section 300 uses by the above-mentioned processing through a network 900 or I/O bus 800 from a host 100, especially relocation information 408, and can carry out a relocation proposal for correction, an addition, deletion, etc.
[0042] By performing the above-mentioned processing, based on the acquired operating condition information and the set-up class attribute, in the storage subsystem 200, physical relocation of a logic field can be performed automatically, and the storage subsystem 200 can be optimized. By repeating the further above-mentioned relocation decision and processing of activation, and correcting arrangement, the optimization error factor of fluctuation of an operating condition or others is absorbable.
[0043] Especially, a user or a customer engineer can perform optimization by relocation simple by the above-mentioned processing. Since a user or a customer engineer can manage storage 500 in the unit of a class 600, it does not need to manage attributes, such as engine performance of storage 500, dependability, and a property, about said each storage 500. Furthermore, a user or a customer engineer can set up the class 600 in which each attribute of storage 500 has the same attribute also to the group which is not equal if needed, and can treat it as one management unit. However, it is able for one storage 500 to consider that one class 600 is constituted, and to process the above-mentioned relocation by making one storage 500 into a management unit.
[0044] Moreover, a user or a customer engineer can perform the above-mentioned relocation automatically in consideration of the description and schedule of the processing (job) performed by the host 100. Generally, I/O accompanying the processing performed with a computing system and this processing is performed in conformity with the schedule created by the user. A user can specify the period of processing, when it has processing to make into the object of optimization especially, a user can specify an interested period, and can make relocation decision able to process to the storage system 200 by processing of relocation in which it explained with the gestalt of this operation, namely, optimization by the above-mentioned relocation can be realized based on the operating-condition information on said period. Moreover, the inclination of the processing performed with a computing system and I/O shows the periodicity for every month and every year day by day in many cases. Especially, periodicity becomes remarkable when processing is processing based on a routine task. Like the above-mentioned case,

especially a user can specify the period which is interested as a candidate for optimization in a period, and can perform optimization by relocation. Moreover, in the relocation executive operation 313, although accompanied by the copy of the contents of storing within the storage system 200, a user is setting up the period when the demand processing engine performance of processing the storage system's 200 being performed by the time of day currently seldom used or the host 100 is low as activation time of day of the relocation executive operation 313, and it can avoid that I/O to the storage system 200 of processing that the demand processing engine performance in a host 100 is high is checked by the copy.
[0045] In addition, storage 500 may be a storage which may have the engine performance, dependability, a property different, respectively, and an attribute different, respectively, and is specifically especially different like a magnetic disk drive, a magnetic tape unit, and semiconductor memory (cache). Moreover, although [the above-mentioned example] the free-space information 407 is described based on a physical field, it may be described based on the logic field (logical address) corresponding to an intact physical field.
[0046] The gestalt of <gestalt of the second operation> book operation explains the relocation decision by application of the disk unit activity ratio as operating condition information, the upper limit of a class 600, and the engine-performance ranking between classes 600.
[0047] Drawing 10 is the block diagram of the computing system in the gestalt of operation of the 2nd of this invention.
[0048] The computer system of the gestalt of this operation comes to have a host 100, the disk array system 201, and a control terminal 700. The computer system in the gestalt of this operation is equivalent to what used the storage subsystem 200 in the gestalt of the 1st operation as the disk array system 201, and made the store 500 the parity group 501.
[0049] The disk array system 201 has a control section 300 and a disk unit 502. A control section 300 is equivalent to the control section 300 in the gestalt of the 1st operation. The disk unit 502 constitutes RAID (disk array) from n sets (n is two or more integers), and calls the group by n sets of these disk units 502 the parity group 501. As a property of RAID, n sets of the disk units 502 contained in one parity group 501 have the relation on the redundancy that the redundancy data generated from the contents of storing of n-1 set of a disk unit 502 are stored in the one remaining sets. Moreover, it has the relation on data storage -- distributed storing of the contents of storing in which n sets of disk units 502 included redundancy data is carried out at n sets of disk units 502 a sake [on a juxtaposition actuation disposition]. Although it can consider from this relation that each parity group 501 is one unit on actuation Since cost, performance characteristics, etc. for redundancy, Number n, etc. to realize differ from each other, In constituting the disk array system 201, also about the disk unit 502 which the array (parity group 501) from which level and Number n differ is made intermingled in many cases, and constitutes the parity group 501 In order to realize optimal cost performance in constituting the disk array system 201 since cost changes with the engine performance, capacity, etc., two or more sorts of disk units 502 from which the engine performance and capacity differ may be used. Therefore, each parity group 501 who builds the disk array system 201 in the gestalt of this operation does not restrict that attributes, such as engine performance, dependability, and a property, are the same, but presupposes that it is different about especially the engine performance.
[0050] An example of the information 400 corresponding to logic/physics in the gestalt of this operation is shown in drawing 11 .
[0051] The logical address is the address which shows the logic field which a host 100 uses by the read/write processing 310. A physical address is the address which shows the field on the disk unit 502 in which data and said redundancy data are actually stored, and consists of the parity group number, each disk unit number, and the address in a disk unit. The parity group number shows each parity group 501. A disk unit number shows each disk unit 502. The address in a disk unit is the address which shows the field within a disk unit 502. Although a control section 300 uses for and processes the information about redundancy data by said read/write processing 310 etc. as actuation of RAID, it does not touch about said processing by explanation of the gestalt of this operation especially here in order to explain the parity group 502 as one

unit on actuation.
[0052] further -- the gestalt of the 1st operation -- the same -- the parity group 501 -- a user -- or it is classified into two or more groups (class 600) as an initial state, and the classification to a class 600 is set as the class configuration information 401. An example of the class configuration information 401 is shown in drawing 12 .
[0053] A class number is a number which shows each class 600. Parity group number shows the number of the parity groups belonging to each class 600. The parity group number shows the parity group number 501 belonging to each class 600. Similarly, the attribute of each class 600 is set as the class attribute information 402. An example of the class attribute information 402 in the gestalt of this operation is shown in drawing 13 .
[0054] A class number is a number which shows each class 600. An activity ratio upper limit is a upper limit which shows the tolerance of the below-mentioned disk activity ratio, and is applied to the parity group 501 to whom a class 600 belongs. Class intersex ability ranking is the engine-performance ranking between classes 600 (the small thing of a figure presupposes that it is highly efficient). Class intersex ability ranking is based on the above-mentioned engine-performance difference in the parity group 501 who constitutes each class 600. About a relocation activation upper limit and immobilization, it mentions later.
[0055] Drawing 14 explains the operating condition information acquisition processing 311 in the gestalt of this operation.
[0056] A control section 300 acquires the time of the disk unit 502 used in the read/write processing 310 like the gestalt of the 1st operation, finds the time per unit time amount (activity ratio), further, about the parity group 501 to whom a disk unit 502 belongs, computes the average of an activity ratio (step 1300), and records an activity ratio average on the logic field operating condition information 403 as a disk unit activity ratio about the logic field used as a read/write object (step 1310). Moreover, a control section 300 asks for the sum of the disk unit activity ratio of all the logic fields corresponding to the parity group 501 (step 1320), and records it on the physical field operating condition information 404 as the parity group's 501 activity ratio (step 1330).
[0057] An example of the logic field operating condition information 403 in the gestalt of this operation and the physical field operating condition information 404 is shown in drawing 1515 and drawing 16 .
[0058] Time shows the time of every sampling period (a fixed period), the logical address shows a logic field, the parity group number shows each parity group, and the disk unit activity ratio and parity group activity ratio of a logic field show the average activity ratio in said sampling period, respectively. The activity ratio of the above disk units 502 is a value which shows the load concerning a disk unit 502, and since the disk unit 502 may serve as an engine-performance bottleneck when an activity ratio is large, the improvement in the engine performance of the disk array system 201 is expectable by lowering an activity ratio by relocation processing.
[0059] Next, drawing 17 R> 7 explains the relocation decision processing 312.
[0060] A control section 300 acquires the parity group 501 belonging to a class 600 from the class configuration information 401 about each class 600 (step 1300). Then, a control section 300 acquires a horizon with reference to the same relocation decision horizon information 405 as the gestalt of the 1st operation, further, about the parity group 501, acquires the parity group activity ratio of the physical field operating condition information 404 on a horizon, and totals (step 1320). Then, a control section 300 acquires the activity ratio upper limit of a class 600 with reference to the class attribute information 402 (step 1330). It is judged that a control section 300 is [relocation of the logic field corresponding to the parity group 501] required since a parity group activity ratio is compared with a class upper limit, and when a parity group activity ratio is larger than a class upper limit reduces the parity group's 501 activity ratio (step 1340).
[0061] Then, with reference to the logic field operating condition information 403 on a horizon, a control section 300 acquires the disk unit activity ratio of the logic field corresponding to each physical field of the parity group 501 judged that relocation is required, totals (step 1350), and it chooses from what has a large disk unit activity ratio as a logic field to rearrange (step 1360). Selection of a logic field subtracts the disk activity ratio of the logic field chosen from the parity

group's 501 activity ratio, and it is performed until it becomes below the activity ratio upper limit of a class 600 (1370). Since it is thought that the effect to the parity group's 501 activity ratio of the logic field where a disk unit activity ratio is large is also large, and its access frequency to the logic field from a host 100 is also large, it is rearranging preferentially the logic field where a disk unit activity ratio's is large, and the effective engine-performance improvement of the disk array system 201 can be expected.
[0062] A control section 300 looks for the physical field used as the relocation place about the selected logic field. A control section 300 acquires the intact physics field of the parity group 501 who belongs to a high performance class with reference to the class configuration information 401 and the same free-space information 407 as the gestalt of the 1st operation with reference to the class attribute information 402 from the class 600 to which the parity group 501 belongs paying attention to the class 600 (high performance class) of a high order [ranking / engine-performance] (step 1380).
[0063] Furthermore, a control section 300 calculates the forecast of the parity group activity ratio at the time of considering as a relocation place about each intact physics field (step 1390). The intact physics field it can be predicted that does not exceed the upper limit set as the high performance class out of an intact physics field when it considers as a relocation place It chooses as a physical field of a relocation place (step 1400), and a selection result is outputted to relocation information 408 like the gestalt of the 1st operation (step 1410). Processing will be ended if it finishes choosing the physical field of a relocation place about all the selected logic fields (step 1420).
[0064] In the gestalt of this operation, in addition to the gestalt of the 1st operation, a control section 300 holds parity group information 409, and computes an activity ratio forecast from parity group information 409, the logic field operating condition information 403, and the physical field operating condition information 404.
[0065] An example of parity group information 409 is shown in drawing 18 R> 8. The parity group number is a number which shows each parity group 501. A RAID configuration shows the level and the number of a disk of RAID which the parity group 501 constitutes, and a redundancy configuration. The disk unit engine performance shows the performance characteristics of the disk unit 502 which constitutes the parity group 501. About immobilization, it mentions later. By making into the parity group 501 of a high performance class the relocation place of the logic field where a disk unit activity ratio is large in the above-mentioned processing, the disk unit time to the same load can be shortened, and the disk unit activity ratio after relocation of a logic field can be controlled.
[0066] although relocation executive operation 313 is performed like the gestalt of the 1st operation, as it is shown in drawing 19 , before a control section 300 performs the copy for relocation -- the class attribute information 402 -- referring to -- the class 600 of a rearranging agency and a relocation place -- a user -- or the relocation activation upper limit set up as initial condition is acquired (step 1500). Furthermore, with reference to the physical field operating condition information 404, the latest parity group activity ratio of the parity group 501 of a rearranging agency and a relocation place is acquired (step 1510), and when the parity group activity ratio is over the relocation activation upper limit in one [at least] class 600 as a result of the comparison, (steps 1520 and 1530) and the relocation executive operation 313 are stopped or postponed (step 1540).
[0067] It can avoid that a load arises further by said copy when the parity group's 501 activity ratio is large, namely, a user's load is expensive by the above-mentioned processing, and the upper limit for evasion can be set as arbitration every class 600.
[0068] By processing as mentioned above, selection of the logic field physically rearranged based on the operating condition of a disk unit 502 and selection of the physical field of a relocation place can be performed based on a class configuration and an attribute, relocation can distribute the load of a disk unit 502, and arrangement for which the activity ratio of the parity group 501 belonging to a class 600 does not exceed the activity ratio upper limit set as each class 600 can be realized. By repeating processing of relocation decision and activation furthermore, and correcting arrangement, fluctuation and the prediction error of an operating condition are

absorbable.
[0069] Although a control section 300 totals with reference to the parity group activity ratio of the physical field operating condition information 404 on a horizon, and the disk unit activity ratio of the logic field of the logic field operating condition information 403 and being used for decision in the relocation decision processing 312 For example, instead of using the average of all the values of a horizon, the method of using the value of m high orders in a horizon is also considered, and the approach using the value of the m-th high order is also considered (m is one or more integers). A user can choose and use only the characteristic part of an operating condition, and can make the relocation decision processing 312 perform by a user enabling it to choose these approaches.
[0070] In the above-mentioned relocation decision processing 312, although [a control section 300] the required parity group 501 of relocation of a logic field is detected about all the classes 600 of the disk array system 201, a control section 300 is good [a control section] about the class 600 to which the fixed attribute is set with reference to the class attribute information 402 also as outside of the object of detection before said detection. Moreover, a control section 300 is good as outside of the object of detection similarly about the parity group 501 by whom the fixed attribute is set up with reference to parity group information 409. Moreover, although [a control section 300] the physical field of a relocation place is chosen from the intact physics field of the parity group 501 belonging to a high performance class, you may make it engine-performance ranking treat the high-order class 600 as a high performance class further as outside of an object about the class 600 to which the fixed attribute is set in the relocation decision processing 312. Moreover, about the parity group 501 by whom the fixed attribute is set up, it is good also as outside of an object. By treating the class 600 or the parity group 501 by whom the fixed attribute is set up as mentioned above, a user can set up the class 600 or the parity group 501 who wants to produce the effect of physical relocation in the automatic above-mentioned relocation processing, and can be taken as the outside of the object of relocation.
[0071] The gestalt of <gestalt of the third operation> book operation explains relocation decision within the same class 600. The computing system in the gestalt of this operation is the same as that of the gestalt of the 2nd operation. However, with the gestalt of this operation, two or more parity groups 501 belong to one class 600. If processing with the gestalt of this operation removes the relocation decision processing 312, it is the same as that of the gestalt of the 2nd operation. Moreover, selection (step 1600) of the logic field rearranged also about the relocation decision processing 312 is the same as that of the gestalt of the 2nd operation.
[0072] Drawing 20 explains selection of the physical field of the relocation place in the relocation decision processing 312 with the gestalt of this operation.
[0073] Although engine-performance ranking chooses the physical field of a relocation place from the high-order class 600 with the gestalt of the 2nd operation from the class 600 to which the physical field of a rearranging agency belongs, it chooses from parity groups 501 other than the rearranging agency of the same class 600 with the gestalt of this operation. A control section 300 acquires the intact physics field of parity groups 501 other than the rearranging agency belonging to the same class 600 with reference to the class configuration information 401 and the free-space information 407 (step 1610). A control section 300 calculates the forecast of the parity group activity ratio at the time of considering as a relocation place about each intact physics field (step 1620). The intact physics field it can be predicted that does not exceed the upper limit set as the same class 600 out of an intact physics field when it considers as a relocation place It chooses as a physical field of a relocation place (step 1630), and a selection result is outputted to relocation information 408 like the gestalt of the 2nd operation (step 1640). Processing will be ended if it finishes choosing the physical field of a relocation place about all the logic fields to rearrange (step 1650).
[0074] The above-mentioned processing can distribute the load of a disk unit 502 in the same class 600. The parity group 501 of the disk array system 201 can apply the above-mentioned art to the configuration which belongs to one class 600 (single class) altogether. Moreover, when it combines with the art explained with the gestalt of the 2nd operation for example, it sets to selection of the intact physics field of a relocation place, and the case where the intact physics

field for the high-order class 600 where engine-performance ranking is more suitable than the class 600 of a rearranging agency is not obtained, and engine-performance ranking can apply to processing in the top class 600. the activity ratio upper limit from which the art in the gestalt of the 2nd operation and the art in the gestalt of this operation differ about each class 600 when it combines with the art explained with the gestalt of the 2nd operation -- you may use -- namely, -- therefore, the class attribute information 402 may have two kinds of activity ratio upper limits, or difference about each class 600.
[0075] In the relocation decision processing 312 with the gestalt of the 2nd operation with the gestalt of <gestalt of the fourth operation> book operation When the intact physics field of a relocation place is not found from the class 600 of a rearranging agency in the class 600 (high performance class) of a high order [ranking / engine-performance] In order to obtain a relocation place, the engine-performance ranking performed explains processing of the relocation from the high performance class to the class 600 (low engine-performance class) of lower order more.
[0076] The computing system in the gestalt of this operation is the same as that of the gestalt of the 2nd operation. Drawing 21 explains the relocation decision processing 312 in the gestalt of this operation.
[0077] A control section 300 acquires the parity group 501 belonging to a high performance class from the class configuration information 401 (step 1700). Then, a control section 300 acquires a horizon with reference to the same relocation decision horizon information 405 as the gestalt of the 1st operation (step 1710), acquires the disk unit activity ratio of the logic field corresponding to each physical field of the parity group 501 with reference to the logic field operating condition information 403 on a horizon (step 1720), and chooses it from what has a small disk unit activity ratio as a logic field rearranged to a low engine-performance class (step 1730). At this time, selection of a logic field is performed as required (step 1740).
[0078] Then, although the physical field used as the relocation place about the selected logic field is chosen from the parity group 501 belonging to a low engine-performance class, the control section 300 of processing of physical field selection of a relocation place is the same as that of processing with the gestalt of the 2nd operation, if the high performance class made into the relocation place in processing explanation with the gestalt of the 2nd operation is read as a low engine-performance class (step 1750). Moreover, processing of others in the gestalt of this operation is the same as processing with the gestalt of the 2nd operation.
[0079] By performing the above-mentioned processing, when the intact physics field of a relocation place is not found in a high performance class in the relocation decision processing 312 with the gestalt of the 2nd operation, relocation of a logic field can be performed from a high performance class in advance of the relocation to a high performance class to a low engine-performance class, and the intact physics field of a relocation place can be prepared for a high performance class. A control section 300 can prepare intact physics field where a repeat line is sufficient for the above-mentioned processing if needed.
[0080] Although the disk time to the same load may increase about relocation and the disk unit activity ratio after relocation of a logic field may increase since the relocation place of a logic field is made into the parity group 501 of a low engine-performance class, the effect of increase can be suppressed to the minimum by making it rearrange from the logic field where a disk activity ratio is small.
[0081] With the gestalt of <gestalt of the fifth operation> book operation, an access classification attribute is prepared in one of the attributes of a class 600, and the relocation decision for carrying out physical relocation of the logic field where a sequential access is notably performed using an access classification attribute, and the logic field where random access is performed notably automatically in other parity groups 501, and separating them is explained.
[0082] The computing system in the gestalt of this operation is shown in drawing 1010 . In addition to explanation with the gestalt of the 2nd operation, with the gestalt of this operation, the following information which a control section 300 holds is used.
[0083] An example of the class attribute information 402 on the gestalt of this operation is

shown in drawing 22 . In this example, when access classification is added to the example in the gestalt of the 2nd operation and the access classification of a class 600 is set up sequentially, for example, it is shown that it is set up that a class 600 is suitable for a sequential access.
[0084] An example of the logic field operating condition information 403 on the gestalt of this operation is shown in drawing 23 . In this example, the rate of a sequential access and the rate of random access are applied to the example in the gestalt of the 2nd operation.
[0085] Furthermore, in addition to the gestalt of the 2nd operation, in the gestalt of this operation, a control section 300 holds the access classification reference-value information 410 and the logic field attribute information 411.
[0086] An example of the access classification reference-value information 410 is shown in drawing 2424 . a user -- or the reference value used for the judgment of the below-mentioned access classification is set to the access classification reference-value information 410 as initial condition. Moreover, an example of the logic field attribute information 411 is shown in drawing 25 . An access classification hint is the access classification which can be expected to be notably carried out about each logic field, and a user sets it up. About immobilization, it mentions later.
[0087] If processing with the gestalt of this operation removes the operating condition information acquisition processing 311 and the relocation decision processing 312, it is the same as that of the gestalt of the second operation.
[0088] Drawing 26 explains the operating condition information acquisition processing 311 in the gestalt of this operation.
[0089] Like the operating condition information acquisition processing 311 with the gestalt of the 2nd operation, a control section 300 computes the disk unit activity ratio about a logic field (steps 1800 and 1810), analyzes the contents of an activity ratio in the read/write processing 310, computes the ratio of a sequential access and random access about an activity ratio (step 1820), and records an activity ratio and an access classification ratio on the logic field operating condition information 403 (step 1830). Moreover, a control section 300 performs calculation of a parity group activity ratio, and record to the physical field operating condition information 404 like the gestalt of the 2nd operation (steps 1840 and 1850).
[0090] In the relocation decision processing 312 in the gestalt of this operation, selection of the logic field to rearrange is the same as that of the gestalt of the 2nd operation (step 1990). Drawing 27 explains selection of the physical field of the relocation place in the relocation decision processing 312.
[0091] A control section 300 acquires the rate of a sequential access about the logic field to rearrange with reference to the logic field use information 403 (step 1910), and compares with the reference value set as the access classification reference-value information 410 (step 1920). When the rate of a sequential access is larger than a reference value, a control section 300 investigates whether with reference to the class attribute information 402, the class 600 (sequential class) to which access classification is set as it is sequential exists (step 1950). When a sequential class exists, a control section 300 acquires the intact physics field of parity groups 501 other than the rearranging agency belonging to a sequential class with reference to the class configuration information 401 and the free-space information 407 (step 1960). Furthermore, a control section 300 calculates the forecast of the parity group activity ratio at the time of considering as a relocation place about each intact physics field (step 1970). The intact physics field it can be predicted that does not exceed the upper limit set as the sequential class out of an intact physics field when it considers as a relocation place It chooses as a physical field of a relocation place (step 1980), and a selection result is outputted to relocation information 408 like the gestalt of the 2nd operation (step 1990). A control section 300 computes an activity ratio forecast from the same parity group information 409 as the gestalt of the 2nd operation, the logic field operating condition information 403 in the gestalt of this operation, and the physical field operating condition information 404.
[0092] In the aforementioned comparison, when the rate of a sequential access is below a reference value, a control section 300 investigates whether with reference to the logic field attribute information 411, it is set up that an access classification hint is sequential about a logic

field (step 1940). When it is set as the access classification hint that it is sequential, a control section 300 investigates the existence of a sequential class like the above (step 1950), and when a sequential class exists, the physical field of a relocation place is chosen from a sequential class (steps 1960-1990).
[0093] In the aforementioned comparison, the rate of a sequential access is said below reference value, and when an access classification hint is not still more sequential, or when a sequential class does not exist, a control section 300 chooses the physical field of a relocation place from classes 600 other than a sequential class like the gestalt of the 2nd operation (step 2000).
[0094] The logic field where a sequential access is notably performed by the above-mentioned processing using the access classification and the activity ratio upper limit which were set as each class 600 as an attribute to mixture of the remarkable sequential access in the same parity group 501 and random access, and the logic field where random access is performed notably can be automatically rearranged in a different parity group 501, it can separate into separation 502, i.e., a different disk unit, and the response engine performance especially to random access can be improved.
[0095] Moreover, in the above-mentioned processing although [a control section 300] automatic separation by relocation is performed paying attention to a sequential access, it is also possible to perform said separation similarly paying attention to random access.
[0096] Supposing a control section 300 does not rearrange a logic field when the fixed attribute is specified as the logic field with reference to the logic field attribute information 411 when the logic field to rearrange is chosen in the above-mentioned relocation decision processing 312, when there is a logic field considered that especially a user does not want to rearrange, a logic field can make into the outside of the object of relocation by setting up a fixed attribute. The processing about the above-mentioned fixed attribute is using the logic field attribute information 411, and can be applied also to the gestalt of the above-mentioned operation.
[0097]
[Effect of the Invention] The user of a storage subsystem or a customer engineer can do simple the activity for performing arrangement optimization by physical relocation of a storage region.

[Translation done.]

* NOTICES *

JPO and NCIPI are not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is the block diagram of the computing system in the gestalt of operation of the 1st of this invention.
[Drawing 2] It is the flow chart of the read/write processing 310 with the gestalt of operation of the 1st of this invention, and the operating condition information acquisition processing 311.
[Drawing 3] It is drawing showing an example of the information 400 corresponding to logic/physics on the gestalt of operation of the 1st of this invention.
[Drawing 4] It is the flow chart of the relocation decision processing 312 with the gestalt of operation of the 1st of this invention.
[Drawing 5] It is drawing showing an example of the relocation decision horizon information 405 on the gestalt of operation of the 1st of this invention.
[Drawing 6] It is drawing showing an example of the relocation information 408 in the gestalt of operation of the 1st of this invention.
[Drawing 7] It is drawing showing an example of the free-space information 407 on the gestalt of operation of the 1st of this invention.
[Drawing 8] It is the flow chart of the relocation executive operation 313 in the gestalt of operation of the 1st of this invention.
[Drawing 9] It is drawing showing an example of the relocation activation time information 406 in the gestalt of operation of the 1st of this invention.
[Drawing 10] It is the block diagram of the computing system of the gestalt of operation of the 2nd of this invention, and the gestalt of the fifth operation.
[Drawing 11] It is drawing showing an example of the information 400 corresponding to logic/physics on the gestalt of operation of the 2nd of this invention.
[Drawing 12] It is drawing showing an example of the class configuration information 401 in the gestalt of operation of the 2nd of this invention.
[Drawing 13] It is drawing showing an example of the class attribute information 402 on the gestalt of operation of the 2nd of this invention.
[Drawing 14] It is the flow chart of the operating condition information acquisition processing 311 with the gestalt of operation of the 2nd of this invention.
[Drawing 15] It is drawing showing an example of the logic field operating condition information 403 on the gestalt of operation of the 2nd of this invention.
[Drawing 16] It is drawing showing an example of the physical field operating condition information 404 on the gestalt of operation of the 2nd of this invention.
[Drawing 17] It is the flow chart of the relocation decision processing 312 with the gestalt of operation of the 2nd of this invention.
[Drawing 18] It is drawing showing an example of parity group information 409 in the gestalt of operation of the 2nd of this invention.
[Drawing 19] It is the flow chart of the relocation executive operation 313 in the gestalt of operation of the 2nd of this invention.
[Drawing 20] It is the flow chart of the relocation decision processing 312 with the gestalt of operation of the 3rd of this invention.

[Drawing 21] It is the flow chart of the relocation decision processing 312 with the gestalt of operation of the 4th of this invention.
[Drawing 22] It is drawing showing an example of the class attribute information 402 on the gestalt of operation of the 5th of this invention.
[Drawing 23] It is drawing showing an example of the logic field operating condition information 403 on the gestalt of operation of the 5th of this invention.
[Drawing 24] It is drawing showing an example of the access classification reference-value information 410 on the gestalt of operation of the 5th of this invention.
[Drawing 25] It is drawing showing an example of the logic field attribute information 411 on the gestalt of operation of the 5th of this invention.
[Drawing 26] It is the flow chart of the operating condition information acquisition processing 311 with the gestalt of operation of the 5th of this invention.
[Drawing 27] It is the flow chart of the relocation decision processing 312 with the gestalt of operation of the 5th of this invention.
[Description of Notations]
100 Host
200 Storage Subsystem
201 Disk Array System
300 Control Section
310 Read/write Processing
311 Operating Condition Information Acquisition Processing
312 Relocation Decision Processing
313 Relocation Executive Operation
400 Information corresponding to Logic/Physics
401 Class Configuration Information
402 Class Attribute Information
403 Logic Field Operating Condition Information
404 Physical Field Operating Condition Information
405 Relocation Decision Horizon Information
406 Relocation Activation Time Information
407 Free-Space Information
408 Relocation Information
409 Parity Group Information
410 Access Classification Reference-Value Information
411 Logic Field Attribute Information
500 Storage
501 Parity Group
502 Disk Unit
600 Class
700 Control Terminal
800 I/O Bus
900 Network

[Translation done.]

(19)日本国特許庁（ＪＰ）　(12) 公開特許公報（A）　(11)特許出願公開番号 特開2001－67187 (P2001－67187A)
(43)公開日 平成13年3月16日(2001.3.16)

(51)Int.Cl.⁷	識別記号	FI	テーマコード゛(参考)
G06F 3/06	301	G06F 3/06 301A	5B065
	540	540	5B082
12/00	501	12/00 501B	

審査請求 未請求 請求項の数10 OL (全 24 頁)

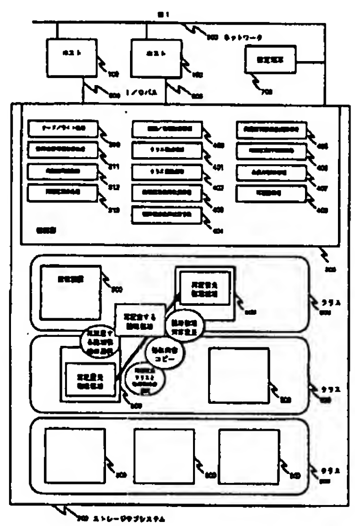
(21)出願番号 特願平11－242713
(22)出願日 平成11年8月30日(1999.8.30)

(71)出願人 000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目6番地
(72)発明者 荒川 敬史
神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内
(72)発明者 茂木 和彦
神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内
(74)代理人 100075096
弁理士 作田 康夫

最終頁に続く

(54)【発明の名称】 ストレージサブシステム及びその制御方法

(57)【要約】
【課題】ストレージサブシステムのユーザまたは保守員が記憶領域の物理的再配置による配置最適化を行うための作業を簡便にするストレージサブシステムおよび制御方法を提供する。
【解決手段】ストレージサブシステム２００は、記憶装置５００を、それぞれ属性を有する複数の組（クラス）６００として管理し、クラス属性に基づき好適な再配置先のクラスを決定する。



1

【特許請求の範囲】
【請求項１】複数の記憶装置と、前記記憶装置の使用状況情報を取得する手段と、前記計算機がリードライト対象とする論理記憶領域と前記記憶装置の第一の物理記憶領域との対応づけを行う手段とを有し、１台以上の計算機に接続するストレージサブシステムの制御方法であって、前記記憶装置は複数の組（クラス）に分類され、前記クラスは設定された属性を有し、前記ストレージサブシステムは、前記使用状況情報および前記クラス属性に基づき前記論理記憶領域に好適な再配置先のクラスを決定し、前記論理記憶領域の再配置先として利用可能な第二の物理記憶領域を前記クラス内から選択し、前記第一の物理記憶領域の内容を前記第二の前記物理記憶領域にコピーするとともに論理記憶領域の対応づけを前記第一の物理記憶領域から前記第二の物理記憶領域へ変更して再配置を行うことを特徴とする前記ストレージサブシステムの制御方法。
【請求項２】請求項１に記載のストレージサブシステムの制御方法であって、ストレージサブシステムは、前記使用状況情報を蓄積し、設定された期間の前記使用状況情報に基づき、論理記憶領域の再配置先を決定し、設定された時間に再配置を行うことを特徴とするストレージサブシステムの制御方法。
【請求項３】請求項１または２に記載のストレージサブシステムの制御方法であって、ストレージサブシステムは、使用状況情報として、記憶装置の単位時間当たりの使用時間（使用率）を用い、各クラスは、属性として設定されたクラス間の性能順位と使用率上限値を有し、前記ストレージサブシステムは、クラスの使用率上限値を超えている記憶装置から再配置する論理記憶領域を選択し、前記論理記憶領域の再配置先のクラスを前記順位の上位のクラスから、各クラスの使用率上限値を超えないように決定することを特徴とするストレージサブシステムの制御方法。
【請求項４】請求項１または２に記載のストレージサブシステムの制御方法であって、ストレージサブシステムは、使用状況情報として、記憶装置の単位時間当たりの使用時間（使用率）を用い、各クラスは、属性として設定されたクラス間の性能順位と使用率上限値を有し、前記ストレージサブシステムは、クラスの使用率上限値を超えている記憶装置から再配置する論理記憶領域を選択し、前記論理記憶領域の再配置先として利用可能な物理記憶領域を同一クラス内の記憶装置から、前記クラスの使用率上限値を超えないように決定することを特徴とするストレージサブシステムの制御方法。
【請求項５】請求項１または２に記載のストレージサブシステムの制御方法であって、ストレージサブシステムは、使用状況情報として、記憶装置の単位時間当たりの使用時間（使用率）を用い、各クラスは属性として設定された対象アクセス種別と使用率上限値を有し、前記ストレージサブシステムは、クラスの使用率上限値を超えている記憶装置から再配置する論理記憶領域を選択し、前記論理記憶領域に対するアクセス種別の分析結果に基づいて前記論理記憶領域の再配置先のクラスを前記対象アクセス種別のクラスから、各クラスの使用率上限値を超えないように決定することを特徴とするストレージサブシステムの制御方法。

2

【請求項６】１台以上の計算機に接続し、複数の記憶装置と、前記記憶装置の使用状況情報を取得する手段と、前記計算機がリードライト対象とする論理記憶領域と前記記憶装置の第一の物理記憶領域との対応づけを行う手段とを有するストレージサブシステムであって、前記複数のディスク装置をそれぞれ属性を有する複数の組（クラス）として管理する手段と、前記使用状況情報および前記クラス属性に基づき前記論理記憶領域に好適な再配置先のクラスを決定する手段と、前記論理記憶領域の再配置先として利用可能な第二の物理記憶領域を前記クラス内から選択する手段と、前記第一の物理記憶領域の内容を前記第二の前記物理記憶領域にコピーするとともに論理記憶領域の対応づけを前記第一の物理記憶領域から前記第二の物理記憶領域へ変更して再配置を行う手段とを有することを特徴とするストレージサブシステム。
【請求項７】請求項６に記載のストレージサブシステムであって、ストレージサブシステムは、前記使用状況情報を蓄積し、設定された期間の前記使用状況情報に基づき、論理記憶領域の再配置先を自動的に決定する手段と、設定された時間に再配置を行う手段とを有することを特徴とするストレージサブシステム。
【請求項８】請求項６または７に記載のストレージサブシステムであって、ストレージサブシステムは、使用状況情報として記憶装置の単位時間当たりの使用時間（使用率）を用いる手段を有し、前記ストレージサブシステムは、各クラスに属性として設定されている使用率上限値を超えている記憶装置から再配置する論理記憶領域を選択する手段と、前記論理記憶領域の再配置先のクラスを各クラスに属性として設定されているクラス間の性能順位から、各クラスの使用率上限値を超えないように決定する手段とを有することを特徴とするストレージサブシステム。
【請求項９】請求項６または７に記載のストレージサブシステムであって、ストレージサブシステムは、使用状況情報として、記憶装置の単位時間当たりの使用時間（使用率）を用いる手段を有し、前記ストレージサブシステムは、属性として設定されたクラスの使用率上限値を超えている記憶装置から再配置する論理記憶領域を選択する手段と、前記論理記憶領域に対するアクセス種別を分析する手段と、対象アクセス種別を属性として設定されたクラスから、前記論理記憶領域の再配置先のクラスを前記分析結果に基づいて各クラスの使用率上限値を超えないように決定する手段とを有することを特徴とす

るストレージサブシステム。

【請求項10】請求項6、7、8、または9に記載のストレージサブシステムであって、ストレージサブシステムは、複数のディスク装置を有するディスクアレイであり、前記ディスク装置の使用率を使用状況情報として用いる手段を有することを特徴とするストレージサブシステム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、複数の記憶装置を有するストレージサブシステム、およびその制御方法に関する。

【0002】

【従来の技術】コンピュータシステムにおいて、高性能を実現する二次記憶システムの1つにディスクアレイシステムがある。ディスクアレイシステムは、複数のディスク装置をアレイ状に配置し、前記各ディスク装置に分割格納されるデータのリード／ライトを、前記各ディスク装置を並列に動作させることによって、高速に行うシステムである。ディスクアレイシステムに関する論文としては、D. A. Patterson, G. Gibson, and R. H. Kats, "A Case for Redundant Arrays of Inexpensive Disks (RAID)" (in Proc. ACM SIGMOD, pp. 109－116, June 1988) がある。この論文では、冗長性を付加したディスクアレイシステムに対し、その構成に応じてレベル1からレベル5の種別を与えている。これらの種別に加えて、冗長性無しのディスクアレイシステムをレベル0と呼ぶこともある。上記の各レベルは冗長性などにより実現するためのコストや性能特性などが異なるため、ディスクアレイシステムを構築するにあたって、複数のレベルのアレイ（ディスク装置の組）を混在させることも多い。ここでは、この組のことをパリティグループと呼ぶ。

【0003】ディスク装置は、性能や容量などによりコストが異なり、ディスクアレイシステムを構築するにあたって最適なコストパフォーマンスを実現するために、やはり性能や容量の異なる複数種のディスク装置を用いることがある。

【0004】ディスクアレイシステムに格納されるデータを上記のようにディスク装置に分散して配置するため、ディスクアレイシステムは、ディスクアレイシステムに接続するホストコンピュータがアクセスする論理記憶領域とディスク装置の記憶領域を示す物理記憶領域の対応づけ（アドレス変換）を行う。特開平9－274544号公報には、ホストコンピュータからの論理記憶領域に対するI／Oアクセスについての情報を取得する手段と、論理記憶領域の物理記憶領域への対応づけを変更して物理的再配置を行う手段により、格納されたデータ

の最適配置を実現するディスクアレイシステムが開示されている。

【0005】

【発明が解決しようとする課題】特開平9－274544号公報に示されるような従来の技術における配置最適化の実行方法については以下の課題がある。

【0006】再配置する論理記憶領域の選択および再配置先の物理記憶領域の選択にあたり、ディスクアレイシステムのユーザまたは保守員が、前記ディスクアレイシステムの構成や個々のディスク装置の特性や性能などの情報を確認して前記選択を行わなければならず、ユーザまたは保守員による作業が煩雑となっていた。

【0007】また、ディスクアレイシステムが選択を自動的に行う場合においても、ユーザまたは保守員が前記個々のディスク装置の情報を確認して選択基準値を規定しなければならず、やはりユーザまたは保守員による作業が煩雑となっていた。特に、上記のように異種のレベルや異種のディスク装置の混在するディスクアレイシステムについては情報管理の煩雑さが増大する。

【0008】また、ディスクアレイシステムが選択のために行うI／Oアクセス情報の参照は、ホストコンピュータおよびディスクアレイシステムを含むシステムで行われる処理のスケジュールの特性を考慮していなかった。一般にコンピュータシステムで行われる処理と処理に伴うI／Oは、ユーザによって作成されたスケジュールに則って行われており、また処理およびI／Oの傾向は日毎、月毎、年毎などの周期性を示す場合も多く、一般にユーザは特定期間の処理およびI／Oに関心があると考えられる。

【0009】また上記従来技術において、再配置による性能チューニング方法については以下の課題がある。物理的再配置による性能チューニング方法は、ディスク装置、すなわち、物理記憶領域の使用状況に変更を加えるものであるが、従来の技術においては、ホストコンピュータからの論理記憶領域に対するI／Oアクセスについての情報を参照するため、再配置する論理記憶領域の選択および再配置先の物理記憶領域の選択にあたり、正しい選択が行えない可能性があった。

【0010】また、ホストコンピュータからのシーケンシャルアクセスとランダムアクセスが顕著に、同一のディスク装置に含まれる別々の物理記憶領域に対して行われる場合でも、シーケンシャルアクセスとランダムアクセスを異なるディスク装置に分離するために、再配置先のディスク装置を任意に特定して自動的再配置を行わせることはできなかった。一般に、ホストコンピュータからの処理要件として、データ長の小さいランダムアクセスには短時間での応答（高応答性能）が求められるが、同一ディスク装置にデータ長の大きいシーケンシャルアクセスが存在する場合、ランダムアクセスの応答時間はシーケンシャルアクセスの処理に阻害されて長くなり、

応答性能は悪化してしまう。

【0011】本発明の第一の目的は、ディスクアレイシステムのユーザまたは保守員が再配置による配置最適化を行うための作業を簡便にすることにある。

【0012】本発明の第二の目的は、ホストコンピュータおよびディスクアレイシステムを含むシステムでの処理のスケジュールを考慮した再配置による配置最適化を可能にすることにある。

【0013】本発明の第三の目的は、再配置する論理記憶領域の選択および再配置先の物理記憶領域の選択にあたり、実際の記憶装置であるディスク装置の使用状況に基づく選択を行う、ディスクアレイシステムの制御方法およびディスクアレイシステムを提供することにある。

【0014】本発明の第四の目的は、ディスクアレイシステムにおける同一ディスク装置での顕著なシーケンシャルアクセスとランダムアクセスの混在に対し、再配置先のディスク装置を任意に特定して再配置によりシーケンシャルアクセスおよびランダムアクセスを異なるディスク装置に自動的に分離することができるようにすることにある。

【0015】

【課題を解決するための手段】上記の第一目的を実現するために、1台以上のホストコンピュータに接続するディスクアレイシステムは、配下の複数のディスク装置の使用状況情報を取得する手段と、ホストコンピュータがリード／ライト対象とする論理記憶領域とディスク装置の第一の物理記憶領域との対応づけを行う手段とを有し、さらに、複数のディスク装置をそれぞれ属性を有する複数の組（クラス）として管理する手段と、使用状況情報およびクラス属性に基づき論理記憶領域に好適な再配置先のクラスを決定する手段と、論理記憶領域の再配置先として利用可能な第二の物理記憶領域をクラス内から選択する手段と、第一の物理記憶領域の内容を前記第二の前記物理記憶領域にコピーするとともに論理記憶領域の対応づけを第一の物理記憶領域から第二の物理記憶領域へ変更して再配置を行う手段を備える。

【0016】また、上記第二の目的を実現するために、ディスクアレイシステムは、使用状況情報を蓄積し、設定された期間の使用状況情報に基づき、論理記憶領域の再配置先を決定する手段と、設定された時間に再配置を行う手段を備えることができる。

【0017】また、上記第三の目的を実現するために、ディスクアレイシステムは、使用状況情報として、ディスク装置の単位時間当たりの使用時間（使用率）を用いる手段を備える。

【0018】また、上記第四の目的を実現するために、ディスクアレイシステムは、各クラスに属性として設定された対象アクセス種別（シーケンシャル／ランダムアクセス種別）と使用率上限値を用いて、クラスの使用率上限値を超えている記憶装置から再配置する論理記憶領

域を選択し、論理記憶領域に対するアクセス種別の分析結果に基づいて論理記憶領域の再配置先のクラスを好適なアクセス種別のクラスから、各クラスの使用率上限値を超えないように決定する手段を備える。

【0019】

【発明の実施の形態】以下、本発明の実施の形態を図1～図27を用いて説明する。

【0020】<第一の実施の形態>本実施の形態では、クラス600に基づく再配置の判断と、再配置判断および実行のスケジューリングについて説明する。

【0021】図1は、本発明の第1の実施の形態における計算機システムの構成図である。

【0022】本実施の形態における計算機システムは、ホスト100、ストレージサブシステム200、制御端末700を有してなる。

【0023】ホスト100は、ストレージサブシステム200にI／Oバス800を介して接続し、ストレージサブシステム200に対しリードやライトのI／Oを行う。I／Oの際、ホスト100は、ストレージサブシステム200の記憶領域について論理領域を指定する。I／Oバス800の例としては、ESCON、SCSI、ファイバチャネルなどがある。

【0024】ストレージサブシステム200は、制御部300および複数の記憶装置500を有する。制御部300は、リード／ライト処理310、使用状況情報取得処理311、再配置判断処理312、及び再配置実行処理313を行う。また、ストレージサブシステム200は、論理／物理対応情報400、クラス構成情報401、クラス属性情報402、論理領域使用状況情報403、物理領域使用状況情報404、再配置判断対象期間情報405、再配置実行時刻情報406、未使用領域情報407、及び再配置情報408を保持する。

【0025】ホスト100、制御部300、および制御端末700は、ネットワーク900で接続される。ネットワーク900の例としては、FDDI、ファイバチャネルなどがある。

【0026】ホスト100、制御部300、および制御端末700には、各々での処理を行うためのメモリ、CPUなど、計算機において一般に用いられる構成要素もそれぞれ存在するが、本実施の形態の説明においては重要でないため、ここでは説明を省略する。

【0027】ホスト100が、ストレージサブシステム200に対してリード／ライトを行う場合のリード／ライト処理310、および使用状況情報取得処理311について図2で説明する。

【0028】リード／ライト処理310において、ホスト100は、ストレージサブシステム200の制御部300に対しリードまたはライトを論理領域を指定して要求する（ステップ1000）。要求を受領した制御部300は、論理／物理対応情報400を用いて論理領域に

対応する物理領域を求め、すなわち論理領域のアドレス（論理アドレス）を物理領域のアドレス（物理アドレス）に変換する（ステップ1010）。続いて制御部300は、リードの場合は、この物理アドレスの記憶装置500からデータを読み出してホスト100に転送し、ライトの場合は、ホスト100から転送されたデータを前記物理アドレスの記憶装置500に格納し（ステップ1020）、さらに後述の使用状況情報取得処理311を行う。リード／ライト要求およびデータ転送はI／Oバス800を介して行われる。

【0029】論理／物理対応情報400の一例を図3に示す。論理アドレスはホスト100がリード／ライト処理310で用いる論理領域を示すアドレスである。物理アドレスは実際にデータが格納される記憶装置500上の領域を示すアドレスであり、記憶装置番号および記憶装置内アドレスからなる。記憶装置番号は個々の記憶装置500を示す。記憶装置内アドレスは記憶装置500内での記憶領域を示すアドレスである。

【0030】次に、使用状況情報取得処理311において制御部300は、リード／ライト処理310においてリード／ライト対象となった論理領域についての論理領域使用状況情報403と、リード／ライト処理310で使用した物理領域についての物理領域使用状況情報404を更新する（ステップ1030、1040）。論理領域使用状況情報403および物理領域使用状況情報404は、例えば使用頻度、使用率、リード／ライトに関する属性など、各々の論理領域と物理領域の各日時の使用状況に関する情報である。論理領域使用状況情報403および物理領域使用状況情報404の具体的な例は、以降の実施の形態で説明する。

【0031】次に、制御部300が行う再配置判断処理312について図4で説明する。

【0032】記憶装置500は、ユーザによって、または初期状態として複数の組（クラス600）に分類されており、クラス600への分類はクラス構成情報401に設定されている。さらに、各クラス600は、ユーザによって、または初期条件として属性を設定されており、属性は、クラス属性情報402に設定されている。クラス属性情報402は、許容使用状況や好適な使用状況やクラス間優先順位などの属性に関する情報である。クラス構成情報401およびクラス属性情報402の具体的な例は、以降の実施の形態で説明する。再配置判断対象期間情報405には、ユーザによってまたは初期条件として再配置判断処理312の対象とする使用状況情報の期間と期間更新情報が設定されている。

【0033】再配置判断対象期間情報405の一例を図5に示す。開始日時から終了日時までの期間が対象期間となる。期間更新情報は次回の対象期間の設定条件であり、例えば毎週、毎日、X時間後などがありうる。制御部300は、対象期間の論理領域使用状況情報403および物理領域使用状況情報404を参照し（ステップ1100）、クラス属性情報402の各クラス600の許容使用状況などと比較して（ステップ1110）、物理的再配置を行うべき論理領域を選択する（ステップ1120）。

【0034】さらに、制御部300は、クラス属性情報402の許容使用状況や好適な使用状況やクラス間優先順位などを参照して（ステップ1130）、論理領域の再配置先のクラス600を選択し（ステップ1140）、さらに、クラス600に属する記憶装置500の中から論理領域の再配置先として未使用の物理領域を選択し（ステップ1150）、選択結果を再配置情報408に出力する（ステップ1160）。

【0035】再配置情報408の一例を図6に示す。論理領域は、再配置する論理領域であり、再配置元物理領域は、論理領域に対応する現在の物理領域を示す記憶装置番号と記憶装置内アドレスであり、再配置先物理領域は、再配置先の物理領域を示す記憶装置番号と記憶装置内アドレスである。図6に示すように再配置の立案は一つ以上行われうる。さらに制御部300は、再配置判断対象期間情報405の期間更新情報を参照して、再配置判断対象期間情報405の対象期間を次回分に更新する（ステップ1170）。上記の処理において制御部300は、論理／物理対応情報400を用い、また前記の未使用の物理領域の検索に未使用領域情報407を用いる。

【0036】未使用領域情報407の一例を図7に示す。記憶装置番号は個々の記憶装置500を示す。記憶装置内アドレスは記憶装置500内での領域を示すアドレスである。記憶装置番号および装置内アドレスは物理領域を示し、使用／未使用の項目は、物理領域の使用／未使用の区別を示す。制御部300は、通常、再配置判断処理312を対象期間以後、後述の再配置実行処理313以前に自動的に行う。

【0037】次に、制御部300が行う再配置実行処理313について図8で説明する。

【0038】再配置実行時刻情報406にはユーザによってまたは初期条件として再配置実行処理313を行う日時と日時更新情報が設定されている。

【0039】再配置実行時刻情報406の一例を図9に示す。制御部300は、設定された日時に以下に説明する再配置実行処理313を自動的に実行する。日時更新情報は次回の再配置実行処理313を行う日時の設定条件であり、例えば毎週、毎日、X時間後などがありうる。制御部300は、再配置情報408に基づき再配置元物理領域に格納している内容を再配置先物理領域にコピーする（ステップ1200）。さらに、コピーが完了して再配置元物理領域の内容が全て再配置先物理領域に反映された時点で、制御部300は、論理／物理対応情報400上の再配置を行う論理領域に対応する物理領域を再配置元物理領域から再配置先物理領域に変更する（ステップ1210）。

【0040】さらに、制御部300は、未使用物理領域470上の再配置先物理領域を使用とし、再配置元物理領域を未使用に変更する（ステップ1220）。さらに制御部300は、再配置実行時刻情報406の日時更新情報を参照して、再配置実行時刻情報406の日時を次回分に更新する（ステップ1230）。

【0041】ユーザまたは保守員は、制御部300が上記の処理で用いている各情報を、制御端末700からネットワーク900を介して、またはホスト100からネットワーク900またはI／Oバス800を介して設定および確認すること、特に、再配置情報408を確認および設定して再配置案を修正や追加や削除などをすることができる。

【0042】上記の処理を行うことによって、取得した使用状況情報および設定されたクラス属性に基づいて、ストレージサブシステム200において論理領域の物理的再配置を自動的に行い、ストレージサブシステム200の最適化を行うことができる。さらに上記の再配置判断および実行の処理を繰り返して配置を修正していくことによって、使用状況の変動やその他の最適化誤差要因を吸収していくことができる。

【0043】特に、上記の処理により、ユーザまたは保守員は再配置による最適化を簡便に行うことができる。ユーザまたは保守員は、記憶装置500をクラス600という単位で管理できるため、記憶装置500の性能や信頼性や特性などの属性を個々の前記記憶装置500について管理する必要はない。さらに、ユーザまたは保守員は、記憶装置500の個々の属性が等しくない組に対しても、必要に応じて同一の属性を持つクラス600を設定して、1つの管理単位として扱うことができる。ただし、1つの記憶装置500が1つのクラス600を構成すると見なして1つの記憶装置500を管理単位として上記の再配置の処理を行うことも可能である。

【0044】また、ユーザまたは保守員は、ホスト100で行われる処理（ジョブ）の特徴やスケジュールを考慮して、上記の再配置を自動的に行うことができる。一般に、計算機システムで行われる処理と、この処理に伴うI／Oは、ユーザによって作成されたスケジュールに則って行われる。ユーザは、特に最適化の対象としたい処理を有する場合、処理の期間を特定することが可能であり、本実施の形態で説明した再配置の処理によって、ユーザは関心のある期間を指定して再配置判断の処理をストレージシステム200に行わせ、すなわち、前記期間の使用状況情報に基づいて上記の再配置による最適化を実現することができる。また、計算機システムで行われる処理およびI／Oの傾向は日毎、月毎、年毎などの周期性を示す場合も多い。特に、処理が定型業務に基づく処理である場合には、周期性が顕著となる。前述の場合と同様にユーザは、周期において特に最適化対象として関心のある期間を指定して再配置による最適化を行うことができる。また、再配置実行処理313では、ストレージシステム200内で格納内容のコピーを伴うが、ユーザはストレージシステム200があまり使用されていない時刻やホスト100で実行されている処理の要求処理性能が低い期間を再配置実行処理313の実行時刻として設定することで、ホスト100での要求処理性能が高い処理のストレージシステム200へのI／Oがコピーにより阻害されることを回避できる。

【0045】なお、記憶装置500は、それぞれ異なる性能、信頼性、特性や属性を持っていてよく、特に具体的には、磁気ディスク装置、磁気テープ装置、半導体メモリ（キャッシュ）のように異なる記憶媒体であってもよい。また、上記の例では未使用領域情報407は物理領域に基づいて記述されているとしたが、未使用の物理領域に対応する論理領域（論理アドレス）に基づいて記述されていてもよい。

【0046】<第二の実施の形態>本実施の形態では、使用状況情報としてのディスク装置使用率の適用と、クラス600の上限値およびクラス600間の性能順位による再配置判断について説明する。

【0047】図10は、本発明の第2の実施の形態における計算機システムの構成図である。

【0048】本実施の形態の計算機システムは、ホスト100、ディスクアレイシステム201、制御端末700を有してなる。本実施の形態における計算機システムは、第1の実施の形態でのストレージサブシステム200をディスクアレイシステム201とし、記憶装置500をパリティグループ501としたものに相当する。

【0049】ディスクアレイシステム201は、制御部300とディスク装置502を有する。制御部300は、第1の実施の形態での制御部300に相当する。ディスク装置502は、n台（nは2以上の整数）でRAID（ディスクアレイ）を構成しており、このn台のディスク装置502による組をパリティグループ501と呼ぶ。RAIDの性質として、1つのパリティグループ501に含まれるn台のディスク装置502は、n－1台のディスク装置502の格納内容から生成される冗長データが残りの1台に格納されるといった冗長性上の関係を持つ。またn台のディスク装置502は、冗長データを含めた格納内容が、並列動作性向上のためにn台のディスク装置502に分散格納されるなど、データ格納上の関係を持つ。この関係から各パリティグループ501を動作上の1単位とみなすことができるが、冗長性や台数nなどにより実現するためのコストや性能特性などが異なるため、ディスクアレイシステム201を構成するにあたって、レベルや台数nの異なるアレイ（パリティグループ501）を混在させることも多く、またパリティグループ501を構成するディスク装置502につ

いても、性能や容量などによりコストが異なるため、ディスクアレイシステム２０１を構成するにあたって最適なコストパフォーマンスを実現するために性能や容量の異なる複数種のディスク装置５０２を用いることもある。よって本実施の形態においてディスクアレイシステム２０１を構築する各パリティグループ５０１は性能、信頼性、特性などの属性が同一であるとは限らず、特に性能について差異があるとする。

【００５０】本実施の形態における論理／物理対応情報４００の一例を図１１に示す。

【００５１】論理アドレスは、ホスト１００がリード／ライト処理３１０で用いる論理領域を示すアドレスである。物理アドレスは実際にデータと前記冗長データが格納されるディスク装置５０２上の領域を示すアドレスであり、パリティグループ番号と各々のディスク装置番号およびディスク装置内アドレスからなる。パリティグループ番号は個々のパリティグループ５０１を示す。ディスク装置番号は個々のディスク装置５０２を示す。ディスク装置内アドレスはディスク装置５０２内での領域を示すアドレスである。制御部３００は、ＲＡＩＤの動作として、冗長データに関する情報を前記リード／ライト処理３１０などで用いて処理するが、本実施の形態の説明では、パリティグループ５０２を動作上の１単位として説明するため、前記処理に関してはここでは特にふれない。

【００５２】さらに第１の実施の形態と同様に、パリティグループ５０１は、ユーザによってまたは初期状態として複数の組（クラス６００）に分類されており、クラス６００への分類はクラス構成情報４０１に設定されている。クラス構成情報４０１の一例を図１２に示す。

【００５３】クラス番号は各クラス６００を示す番号である。パリティグループ数は各クラス６００に属するパリティグループの数を示す。パリティグループ番号は各クラス６００に属するパリティグループ番号５０１を示す。同様に、各クラス６００の属性は、クラス属性情報４０２に設定されている。本実施の形態におけるクラス属性情報４０２の一例を図１３に示す。

【００５４】クラス番号は、各クラス６００を示す番号である。使用率上限値は後述のディスク使用率の許容範囲を示す上限値であり、クラス６００の属するパリティグループ５０１に適用する。クラス間性能順位は、クラス６００間の性能順位（数字の小さいものが高性能とする）である。クラス間性能順位は各クラス６００を構成するパリティグループ５０１の前述の性能差異に基づく。再配置実行上限値および固定については後述する。

【００５５】本実施の形態における使用状況情報取得処理３１１について図１４で説明する。

【００５６】制御部３００は、第１の実施の形態と同様に、リード／ライト処理３１０において使用したディスク装置５０２の使用時間を取得して単位時間当たりの使用時間（使用率）を求め、さらに、ディスク装置５０２が属するパリティグループ５０１について、使用率の平均を算出し（ステップ１３００）、使用率平均を、リード／ライト対象となった論理領域についてのディスク装置使用率として論理領域使用状況情報４０３に記録する（ステップ１３１０）。また制御部３００は、パリティグループ５０１に対応する全論理領域のディスク装置使用率の和を求め（ステップ１３２０）、パリティグループ５０１の使用率として物理領域使用状況情報４０４に記録する（ステップ１３３０）。

【００５７】本実施の形態における論理領域使用状況情報４０３および物理領域使用状況情報４０４の一例を図１５および図１６に示す。

【００５８】日時はサンプリング間隔（一定期間）毎の日時を示し、論理アドレスは論理領域を示し、パリティグループ番号は個々のパリティグループを示し、論理領域のディスク装置使用率およびパリティグループ使用率はそれぞれ前記サンプリング間隔での平均使用率を示す。上記のようなディスク装置５０２の使用率はディスク装置５０２にかかる負荷を示す値であり、使用率が大きい場合は、ディスク装置５０２が性能ボトルネックとなっている可能性があるため、再配置処理で使用率を下げることによりディスクアレイシステム２０１の性能向上が期待できる。

【００５９】次に、再配置判断処理３１２について図１７で説明する。

【００６０】制御部３００は、各クラス６００について、クラス６００に属するパリティグループ５０１をクラス構成情報４０１から取得する（ステップ１３００）。続いて、制御部３００は、第１の実施の形態と同様の再配置判断対象期間情報４０５を参照して対象期間を取得し、さらにパリティグループ５０１について、対象期間の物理領域使用状況情報４０４のパリティグループ使用率を取得し集計する（ステップ１３２０）。続いて、制御部３００は、クラス属性情報４０２を参照してクラス６００の使用率上限値を取得する（ステップ１３３０）。制御部３００は、パリティグループ使用率とクラス上限値を比較し、パリティグループ使用率がクラス上限値より大きい場合は、パリティグループ５０１の使用率を減らすために、パリティグループ５０１に対応する論理領域の再配置が必要と判断する（ステップ１３４０）。

【００６１】続いて、制御部３００は、対象期間の論理領域使用状況情報４０３を参照して、再配置が必要と判断したパリティグループ５０１の各物理領域に対応する論理領域のディスク装置使用率を取得し集計して（ステップ１３５０）、ディスク装置使用率の大きいものから、再配置する論理領域として選択する（ステップ１３６０）。論理領域の選択は、パリティグループ５０１の使用率から選択した論理領域のディスク使用率を減算していき、クラス６００の使用率上限値以下になるまで行う（１３７０）。ディスク装置使用率の大きい論理領域は、パリティグループ５０１の使用率に対する影響も大きく、またホスト１００からの論理領域に対するアクセス頻度も大きいと考えられるため、ディスク装置使用率の大きい論理領域を優先的に再配置することで、ディスクアレイシステム２０１の効果的な性能改善が期待できる。

【００６２】制御部３００は、選択された論理領域についての再配置先となる物理領域を探す。制御部３００は、クラス属性情報４０２を参照し、パリティグループ５０１の属するクラス６００より性能順位が高位のクラス６００（高性能クラス）に注目し、クラス構成情報４０１および第１の実施の形態と同様の未使用領域情報４０７を参照して高性能クラスに属するパリティグループ５０１の未使用物理領域を取得する（ステップ１３８０）。

【００６３】さらに、制御部３００は、各未使用物理領域について、再配置先とした場合のパリティグループ使用率の予測値を求め（ステップ１３９０）、未使用物理領域の中から、再配置先とした場合に高性能クラスに設定されている上限値を超えないと予測できる未使用物理領域を、再配置先の物理領域として選択し（ステップ１４００）、選択結果を第１の実施の形態と同様に、再配置情報４０８に出力する（ステップ１４１０）。選択した全ての論理領域について再配置先の物理領域を選択し終えたら処理を終了する（ステップ１４２０）。

【００６４】本実施の形態において、制御部３００は、第１の実施の形態に加えてパリティグループ情報４０９を保持し、パリティグループ情報４０９、論理領域使用状況情報４０３、及び物理領域使用状況情報４０４から使用率予測値を算出する。

【００６５】パリティグループ情報４０９の一例を図１８に示す。パリティグループ番号は個々のパリティグループ５０１を示す番号である。ＲＡＩＤ構成はパリティグループ５０１が構成するＲＡＩＤのレベルやディスク台数や冗長度構成を示す。ディスク装置性能はパリティグループ５０１を構成するディスク装置５０２の性能特性を示す。固定については後述する。上記の処理においてディスク装置使用率の大きい論理領域の再配置先を高性能クラスのパリティグループ５０１とすることで、同一負荷に対するディスク装置使用時間を短縮でき、論理領域の再配置後のディスク装置使用率を抑制できる。

【００６６】再配置実行処理３１３は、第１の実施の形態と同様に行われるが、図１９に示すように、制御部３００は、再配置のためのコピーを行う前にクラス属性情報４０２を参照し、再配置元および再配置先のクラス６００について、ユーザによってまたは初期条件として設定された再配置実行上限値を取得する（ステップ１５００）。さらに物理領域使用状況情報４０４を参照して、再配置元および再配置先のパリティグループ５０１の直近のパリティグループ使用率を取得し（ステップ１５１０）、比較の結果少なくとも一方のクラス６００においてパリティグループ使用率が再配置実行上限値を超えていた場合は（ステップ１５２０、１５３０）、再配置実行処理３１３を中止または延期する（ステップ１５４０）。

【００６７】上記処理によりユーザは、パリティグループ５０１の使用率が大きくすなわち負荷が高い場合に前記コピーによりさらに負荷が生じることを回避することができ、また回避のための上限値をクラス６００毎に任意に設定することができる。

【００６８】上記のように処理することによって、ディスク装置５０２の使用状況に基づいて物理的に再配置する論理領域の選択、および再配置先の物理領域の選択を、クラス構成および属性に基づいて行い、再配置によりディスク装置５０２の負荷を分散して、各クラス６００に設定されている使用率上限値を、クラス６００に属するパリティグループ５０１の使用率が超えない配置を実現することができる。さらに再配置判断および実行の処理を繰り返して配置を修正していくことによって、使用状況の変動や予測誤差を吸収していくことができる。

【００６９】再配置判断処理３１２において、制御部３００は、対象期間の物理領域使用状況情報４０４のパリティグループ使用率や、論理領域使用状況情報４０３の論理領域のディスク装置使用率を参照して集計し、判断に用いるとしたが、例えば、対象期間の全ての値の平均を用いる代わりに、対象期間中の上位m個の値を用いる方法も考えられ、また上位m番目の値を用いる方法も考えられる（mは１以上の整数）。これらの方法をユーザが選択できるようにすることで、ユーザは使用状況の特徴的な部分のみを選択して用い、再配置判断処理３１２を行わせることができる。

【００７０】上記の再配置判断処理３１２において、制御部３００は、ディスクアレイシステム２０１の全てのクラス６００について、論理領域の再配置の必要なパリティグループ５０１の検出を行うとしたが、前記検出の前に制御部３００がクラス属性情報４０２を参照し、固定属性が設定されているクラス６００については、検出の対象外としてもよい。また同様に、制御部３００がパリティグループ情報４０９を参照し、固定属性が設定されているパリティグループ５０１については検出の対象外としてもよい。また、再配置判断処理３１２において、制御部３００は、高性能クラスに属するパリティグループ５０１の未使用物理領域から再配置先の物理領域を選択するとしたが、固定属性が設定されているクラス６００については対象外として、さらに性能順位が高位のクラス６００を高性能クラスとして扱うようにしてもよい。また固定属性が設定されているパリティグループ５０１については対象外としてもよい。上記のように固

定属性が設定されているクラス６００またはパリティグループ５０１を扱うことによって、ユーザは上記の自動的な再配置処理において物理的な再配置の影響を生じさせたくないクラス６００またはパリティグループ５０１を設定し、再配置の対象外とすることができる。

【００７１】＜第三の実施の形態＞本実施の形態では、同一クラス６００内での再配置判断について説明する。本実施の形態での計算機システムは、第２の実施の形態と同様である。ただし、本実施の形態では１つのクラス６００に複数のパリティグループ５０１が属する。本実施の形態での処理は、再配置判断処理３１２を除いては第２の実施の形態と同様である。また、再配置判断処理３１２についても、再配置する論理領域の選択（ステップ１６００）は、第２の実施の形態と同様である。

【００７２】本実施の形態での再配置判断処理３１２における、再配置先の物理領域の選択について図２０で説明する。

【００７３】第２の実施の形態では再配置先の物理領域を再配置元の物理領域の属するクラス６００より性能順位が高位のクラス６００から選択するが、本実施の形態では同一クラス６００の再配置元以外のパリティグループ５０１から選択する。制御部３００は、クラス構成情報４０１と未使用領域情報４０７を参照して、同一クラス６００に属する再配置元以外のパリティグループ５０１の未使用物理領域を取得する（ステップ１６１０）。制御部３００は、各未使用物理領域について、再配置先とした場合のパリティグループ使用率の予測値を求め（ステップ１６２０）、未使用物理領域の中から、再配置先とした場合に同一クラス６００に設定されている上限値を超えないと予測できる未使用物理領域を、再配置先の物理領域として選択し（ステップ１６３０）、選択結果を第２の実施の形態同様に、再配置情報４０８に出力する（ステップ１６４０）。再配置する全ての論理領域について再配置先の物理領域を選択し終えたら処理を終了する（ステップ１６５０）。

【００７４】上記の処理により、同一クラス６００内においてディスク装置５０２の負荷を分散することができる。上記の処理方法は例えばディスクアレイシステム２０１のパリティグループ５０１が全て１つのクラス６００（単一クラス）に属する構成に適用することができる。また、例えば、第２の実施の形態で説明した処理方法と組み合わせた場合に、再配置先の未使用物理領域の選択において、再配置元のクラス６００より性能順位が高位のクラス６００に適当な未使用物理領域が得られなかった場合や、性能順位が最上位のクラス６００での処理に適用できる。第２の実施の形態で説明した処理方法と組み合わせた場合は、第２の実施の形態での処理方法と本実施の形態での処理方法とが各クラス６００について異なる使用率上限値を用いてもよく、すなわち、そのためにクラス属性情報４０２が各クラス６００について二種類の使用率上限値または差分を有してもよい。

【００７５】＜第四の実施の形態＞本実施の形態では、第２の実施の形態での再配置判断処理３１２において、再配置元のクラス６００より性能順位が高位のクラス６００（高性能クラス）に再配置先の未使用物理領域が見つからなかった場合に、再配置先を得るために先立って行われる、性能順位がより低位のクラス６００（低性能クラス）への高性能クラスからの再配置の処理について説明する。

【００７６】本実施の形態での計算機システムは、第２の実施の形態と同様である。本実施の形態における再配置判断処理３１２について図２１で説明する。

【００７７】制御部３００は、高性能クラスに属するパリティグループ５０１をクラス構成情報４０１から取得する（ステップ１７００）。続いて制御部３００は、第１の実施の形態と同様の再配置判断対象期間情報４０５を参照して対象期間を取得し（ステップ１７１０）、対象期間の論理領域使用状況情報４０３を参照して、パリティグループ５０１の各物理領域に対応する論理領域のディスク装置使用率を取得し（ステップ１７２０）、ディスク装置使用率の小さいものから、低性能クラスへ再配置する論理領域として選択する（ステップ１７３０）。このとき論理領域の選択は必要なだけ行われる（ステップ１７４０）。

【００７８】続いて制御部３００は、選択された論理領域についての再配置先となる物理領域を、低性能クラスに属するパリティグループ５０１から選択するが、再配置先の物理領域選択の処理は、第２の実施の形態での処理説明において再配置先としている高性能クラスを低性能クラスと読み替えれば、第２の実施の形態での処理と同様である（ステップ１７５０）。また、本実施の形態におけるその他の処理も第２の実施の形態での処理と同様である。

【００７９】上記の処理を行うことで、第２の実施の形態での再配置判断処理３１２において高性能クラスに再配置先の未使用物理領域が見つからなかった場合に、高性能クラスから低性能クラスへ論理領域の再配置を、高性能クラスへの再配置に先立って行い、再配置先の未使用物理領域を高性能クラスに用意することができる。制御部３００は、上記の処理を必要に応じ繰り返し行って、十分な未使用物理領域を用意することができる。

【００８０】論理領域の再配置先を低性能クラスのパリティグループ５０１とするため、同一負荷に対するディスク使用時間が再配置について増大し、論理領域の再配置後のディスク装置使用率が増大する可能性があるが、ディスク使用率の小さい論理領域から再配置していくようにすることで、増大の影響を最小限に抑えることができる。

【００８１】＜第五の実施の形態＞本実施の形態では、クラス６００の属性の１つにアクセス種別属性を設け、アクセス種別属性を用いてシーケンシャルアクセスが顕著に行われる論理領域とランダムアクセスが顕著に行われる論理領域とを、他のパリティグループ５０１に自動的に物理的再配置して分離するための再配置判断について説明する。

【００８２】本実施の形態における計算機システムは図１０に示したものである。本実施の形態では、第２の実施の形態での説明に加え、制御部３００が保持する下記の情報を用いる。

【００８３】本実施の形態でのクラス属性情報４０２の一例を図２２に示す。この例では、第２の実施の形態での例に対しアクセス種別が加えられており、クラス６００のアクセス種別が、例えばシーケンシャルに設定されている場合は、クラス６００がシーケンシャルアクセスに好適であると設定されていることを示す。

【００８４】本実施の形態での論理領域使用状況情報４０３の一例を図２３に示す。この例では、第２の実施の形態での例に対し、シーケンシャルアクセス率およびランダムアクセス率が加えられている。

【００８５】さらに、本実施の形態において制御部３００は、第２の実施の形態に加え、アクセス種別基準値情報４１０と論理領域属性情報４１１を保持する。

【００８６】アクセス種別基準値情報４１０の一例を図２４に示す。ユーザによりまたは初期条件として、アクセス種別基準値情報４１０には後述のアクセス種別の判定に用いる基準値が設定されている。また、論理領域属性情報４１１の一例を図２５に示す。アクセス種別ヒントは、各論理領域について顕著に行われると期待できるアクセス種別であり、ユーザが設定する。固定については後述する。

【００８７】本実施の形態での処理は、使用状況情報取得処理３１１および再配置判断処理３１２を除いては第二の実施の形態と同様である。

【００８８】本実施の形態における使用状況情報取得処理３１１について図２６で説明する。

【００８９】制御部３００は、第２の実施の形態での使用状況情報取得処理３１１と同様に、論理領域についてのディスク装置使用率を算出し（ステップ１８００、１８１０）、リード／ライト処理３１０での使用率内容を分析して、使用率についてシーケンシャルアクセスとランダムアクセスの比率を算出し（ステップ１８２０）、使用率およびアクセス種別比率を論理領域使用状況情報４０３に記録する（ステップ１８３０）。また、制御部３００は、第２の実施の形態と同様にパリティグループ使用率の算出と物理領域使用状況情報４０４への記録を行う（ステップ１８４０、１８５０）。

【００９０】本実施の形態における再配置判断処理３１２において、再配置する論理領域の選択は第２の実施の形態と同様である（ステップ１９９０）。再配置判断処理３１２での再配置先の物理領域の選択について図２７で説明する。

【００９１】制御部３００は、論理領域使用情報４０３を参照し、再配置する論理領域についてのシーケンシャルアクセス率を取得し（ステップ１９１０）、アクセス種別基準値情報４１０に設定されている基準値と比較する（ステップ１９２０）。シーケンシャルアクセス率が基準値より大きい場合、制御部３００は、クラス属性情報４０２を参照し、アクセス種別がシーケンシャルと設定されているクラス６００（シーケンシャルクラス）が存在するか調べる（ステップ１９５０）。シーケンシャルクラスが存在する場合、制御部３００は、クラス構成情報４０１と未使用領域情報４０７を参照して、シーケンシャルクラスに属する再配置元以外のパリティグループ５０１の未使用物理領域を取得する（ステップ１９６０）。さらに制御部３００は、各未使用物理領域について、再配置先とした場合のパリティグループ使用率の予測値を求め（ステップ１９７０）、未使用物理領域の中から、再配置先とした場合にシーケンシャルクラスに設定されている上限値を超えないと予測できる未使用物理領域を、再配置先の物理領域として選択し（ステップ１９８０）、選択結果を第２の実施の形態同様に再配置情報４０８に出力する（ステップ１９９０）。制御部３００は、使用率予測値を、第２の実施の形態と同様のパリティグループ情報４０９と本実施の形態における論理領域使用状況情報４０３および物理領域使用状況情報４０４から算出する。

【００９２】前記の比較において、シーケンシャルアクセス率が基準値以下である場合、制御部３００は、論理領域属性情報４１１を参照し、論理領域についてアクセス種別ヒントがシーケンシャルと設定されているか調べる（ステップ１９４０）。アクセス種別ヒントにシーケンシャルと設定されていた場合、上記と同様に制御部３００は、シーケンシャルクラスの有無を調べ（ステップ１９５０）、シーケンシャルクラスが存在する場合は、シーケンシャルクラスから再配置先の物理領域を選択する（ステップ１９６０～１９９０）。

【００９３】前記の比較において、シーケンシャルアクセス率が前記基準値以下であり、さらにアクセス種別ヒントがシーケンシャルでなかった場合、またはシーケンシャルクラスが存在しなかった場合、制御部３００は、第２の実施の形態と同様に、シーケンシャルクラス以外のクラス６００から再配置先の物理領域を選択する（ステップ２０００）。

【００９４】上記の処理により、同一パリティグループ５０１での顕著なシーケンシャルアクセスとランダムアクセスの混在に対し、各クラス６００に属性として設定されたアクセス種別と使用率上限値を用いて、シーケンシャルアクセスが顕著に行われる論理領域とランダムアクセスが顕著に行われる論理領域とを、異なるパリティグループ５０１に自動的に再配置して分離、すなわち異

なるディスク装置５０２に分離することができ、特にランダムアクセスに対する応答性能を改善することができる。

【００９５】また、上記の処理においては制御部３００は、シーケンシャルアクセスに注目して再配置による自動的分離を行うとしたが、同様にランダムアクセスに注目して前記分離を行うことも可能である。

【００９６】上記の再配置判断処理３１２において、再配置する論理領域を選択した時点で、制御部３００が論理領域属性情報４１１を参照し、論理領域に固定属性が指定されている場合は、論理領域を再配置しないとすれば、ユーザが特に再配置を行いたくないと考える論理領域がある場合、固定属性を設定することで論理領域を再配置の対象外とすることができる。上記の固定属性に関する処理は論理領域属性情報４１１を用いることで、前述の実施の形態にも適用できる。

【００９７】

【発明の効果】ストレージサブシステムのユーザ、または保守員が、記憶領域の物理的再配置による配置最適化を行うための作業を簡便にすることができる。

【図面の簡単な説明】

【図１】本発明の第１の実施の形態での計算機システムの構成図である。

【図２】本発明の第１の実施の形態でのリード／ライト処理３１０および使用状況情報取得処理３１１のフローチャートである。

【図３】本発明の第１の実施の形態での論理／物理対応情報４００の一例を示す図である。

【図４】本発明の第１の実施の形態での再配置判断処理３１２のフローチャートである。

【図５】本発明の第１の実施の形態での再配置判断対象期間情報４０５の一例を示す図である。

【図６】本発明の第１の実施の形態での再配置情報４０８の一例を示す図である。

【図７】本発明の第１の実施の形態での未使用領域情報４０７の一例を示す図である。

【図８】本発明の第１の実施の形態での再配置実行処理３１３のフローチャートである。

【図９】本発明の第１の実施の形態での再配置実行時刻情報４０６の一例を示す図である。

【図１０】本発明の第２の実施の形態および第五の実施の形態の計算機システムの構成図である。

【図１１】本発明の第２の実施の形態での論理／物理対応情報４００の一例を示す図である。

【図１２】本発明の第２の実施の形態でのクラス構成情報４０１の一例を示す図である。

【図１３】本発明の第２の実施の形態でのクラス属性情報４０２の一例を示す図である。

【図１４】本発明の第２の実施の形態での使用状況情報取得処理３１１のフローチャートである。

【図１５】本発明の第２の実施の形態での論理領域使用状況情報４０３の一例を示す図である。

【図１６】本発明の第２の実施の形態での物理領域使用状況情報４０４の一例を示す図である。

【図１７】本発明の第２の実施の形態での再配置判断処理３１２のフローチャートである。

【図１８】本発明の第２の実施の形態でのパリティグループ情報４０９の一例を示す図である。

【図１９】本発明の第２の実施の形態での再配置実行処理３１３のフローチャートである。

【図２０】本発明の第３の実施の形態での再配置判断処理３１２のフローチャートである。

【図２１】本発明の第４の実施の形態での再配置判断処理３１２のフローチャートである。

【図２２】本発明の第５の実施の形態でのクラス属性情報４０２の一例を示す図である。

【図２３】本発明の第５の実施の形態での論理領域使用状況情報４０３の一例を示す図である。

【図２４】本発明の第５の実施の形態でのアクセス種別基準値情報４１０の一例を示す図である。

【図２５】本発明の第５の実施の形態での論理領域属性情報４１１の一例を示す図である。

【図２６】本発明の第５の実施の形態での使用状況情報取得処理３１１のフローチャートである。

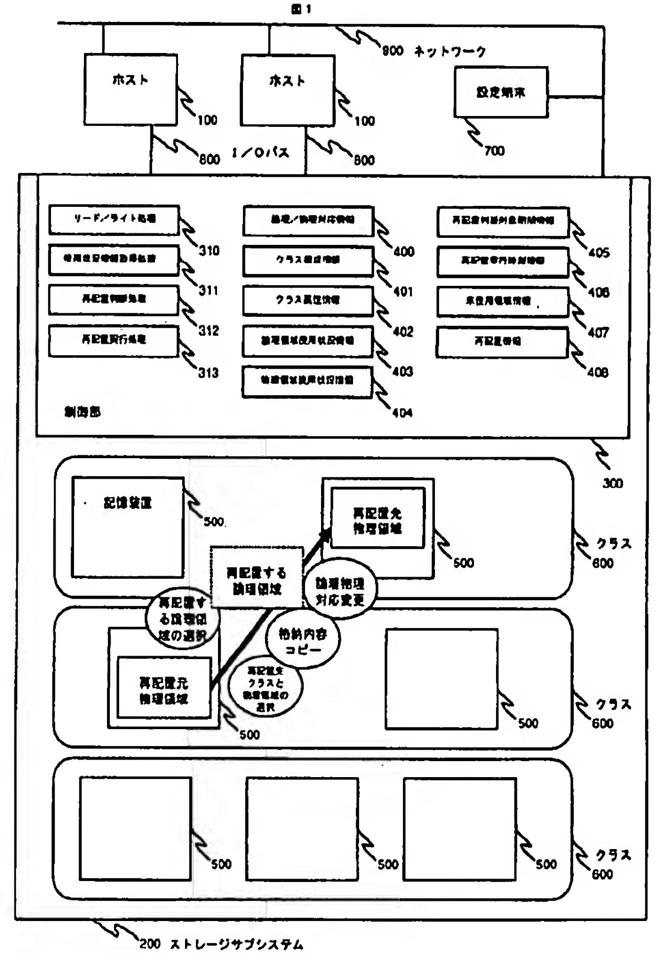
【図２７】本発明の第５の実施の形態での再配置判断処理３１２のフローチャートである。

【符号の説明】

１００　ホスト
２００　ストレージサブシステム
２０１　ディスクアレイシステム
３００　制御部
３１０　リード／ライト処理
３１１　使用状況情報取得処理
３１２　再配置判断処理
３１３　再配置実行処理
４００　論理／物理対応情報
４０１　クラス構成情報
４０２　クラス属性情報
４０３　論理領域使用状況情報
４０４　物理領域使用状況情報
４０５　再配置判断対象期間情報
４０６　再配置実行時刻情報
４０７　未使用領域情報
４０８　再配置情報
４０９　パリティグループ情報
４１０　アクセス種別基準値情報
４１１　論理領域属性情報
５００　記憶装置
５０１　パリティグループ
５０２　ディスク装置

６００　クラス
７００　制御端末
８００　Ｉ／Ｏバス
９００　ネットワーク

【図１】



【図９】

図９

日時	１９９９年８月１１日　２２時０分
日時更新情報	毎日（＋２４時間）

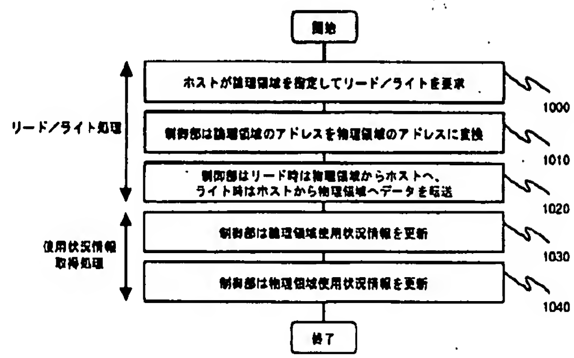
【図２４】

図２４

アクセス判定基準値（％）
７５

【図2】

図2



【図3】

図3

論理アドレス	物理アドレス	
	記憶装置番号	記憶装置内アドレス
0～999	0	0～999
1000～1999	0	1000～1999
2000～2999	1	0～999
3000～3999	1	1000～1999

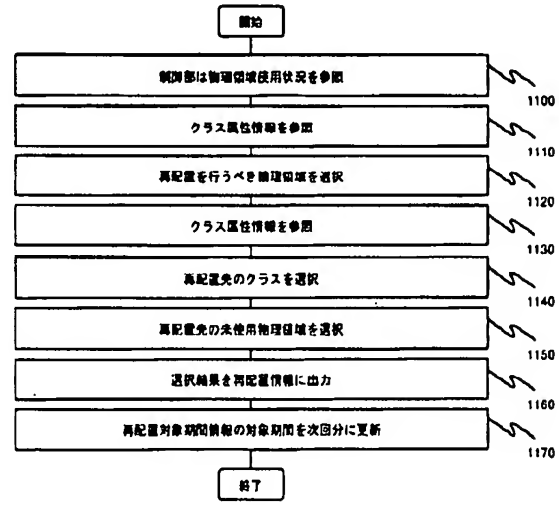
【図5】

図5

開始日時	1999年8月11日　8時30分
終了日時	1999年8月11日　17時15分
期間更新情報	毎日（＋24時間）

【図4】

図4



【図6】

図6

番号	論理領域	再配置元物理領域		再配置先物理領域	
		記憶装置番号	記憶装置内アドレス	記憶装置番号	記憶装置内アドレス
1	0～999	0	0～999	10	0～999
2	1000～1999	0	1000～1999	10	1000～1999

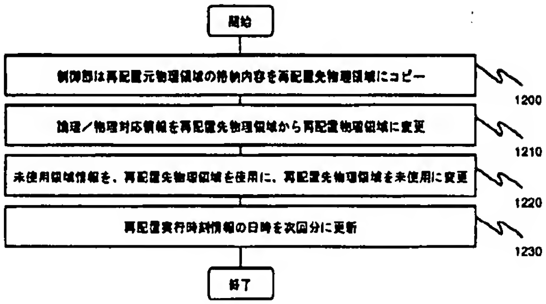
【図7】

図7

記憶装置番号	記憶装置内アドレス	使用／未使用
0	0～999	使用
0	1000～1999	使用
0	2000～2999	未使用
0	3000～3999	未使用

【図８】

図８



【図１１】

図１１

論理アドレス	物理アドレス				
	パリティグループ番号	データ		冗長データ	
		記憶装置番号	記憶装置内アドレス	記憶装置番号	記憶装置内アドレス
0～999	100	0	0～999	20	0～999
1000～1999	100	0	1000～1999	20	1000～1999
2000～2999	101	1	0～999	41	0～999
3000～3999	101	1	1000～1999	41	1000～1999

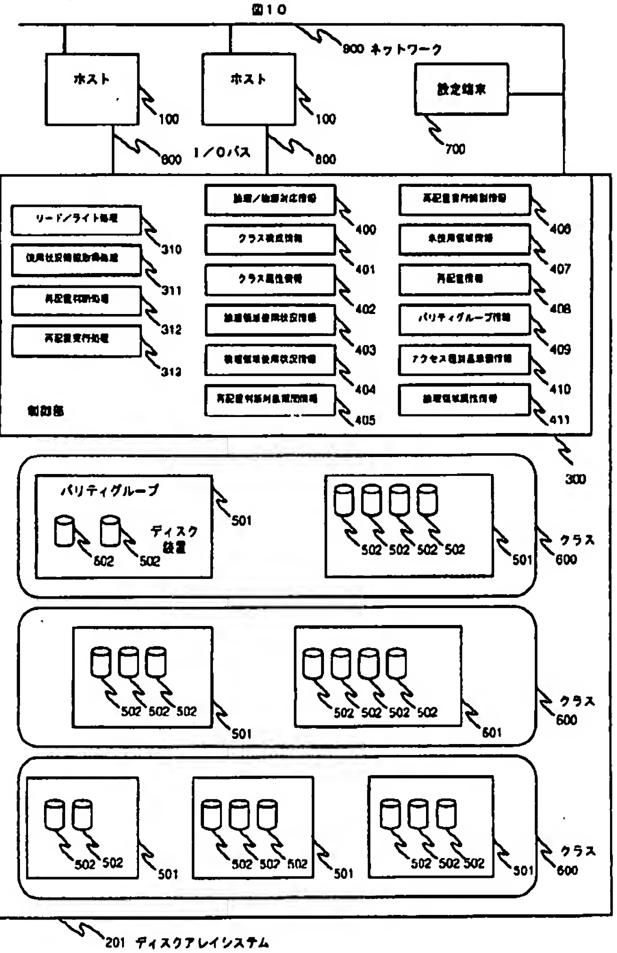
【図１２】

図１２

クラス番号	パリティグループ数	パリティグループ番号
0	3	100, 110, 120
1	2	101, 111
2	4	102, 112, 122, 132

【図１０】

図１０



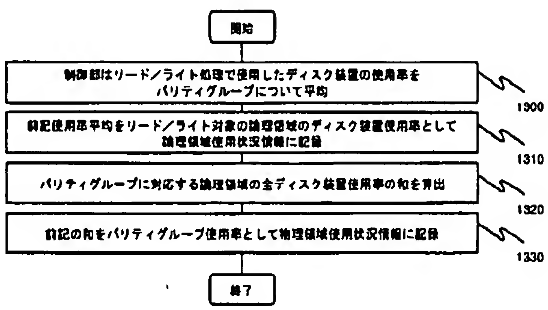
【図13】

図13

クラス番号	使用率上限値（%）	クラス間性能順位	再配置実行上限値（%）	固定
0	60	1	70	−
1	70	2	80	固定
2	80	3	90	−

【図14】

図14



【図15】

図15

日時	論理アドレス	ディスク装置使用率（%）
1999年8月11日 8時0分	0〜999	18
	1000〜1999	32
1999年8月11日 8時15分	0〜999	20
	1000〜1999	30
1999年8月11日 8時30分	0〜999	22
	1000〜1999	28

【図16】

図16

日時	パリティグループ番号	使用率（%）
1999年8月11日 8時0分	100	68
	101	52
1999年8月11日 8時15分	100	70
	101	50
1999年8月11日 8時30分	100	72
	101	48

【図18】

図18

パリティグループ番号	RAID構成	ディスク装置性能	固定
100	RAID5 3D1P	110	−
101	RAID1 1D1P	100	固定
102	RAID5 6D1P	95	−

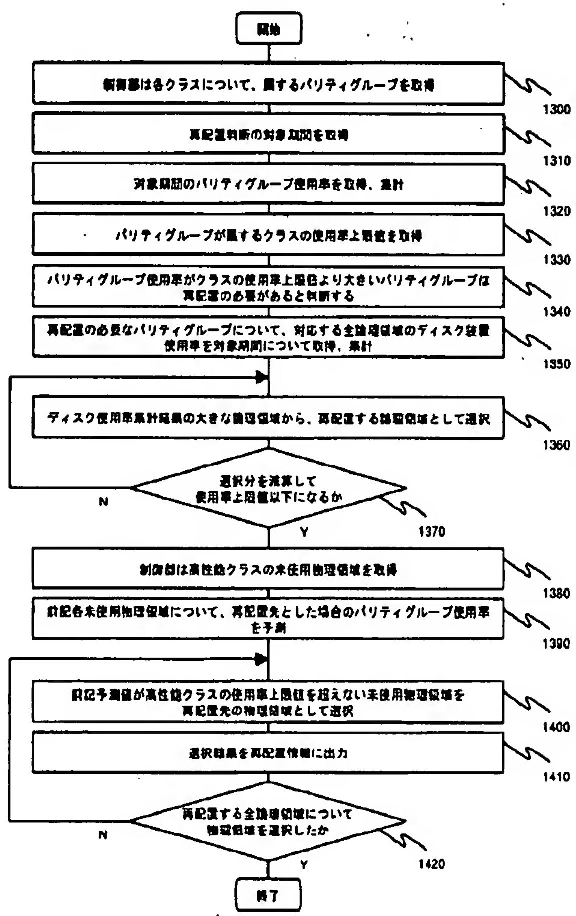
【図22】

図22

クラス番号	使用率上限値（%）	クラス間性能順位	再配置実行上限値（%）	固定	アクセス種別
0	60	1	70	−	−
1	70	2	80	−	−
2	80	3	90	−	シーケンシャル

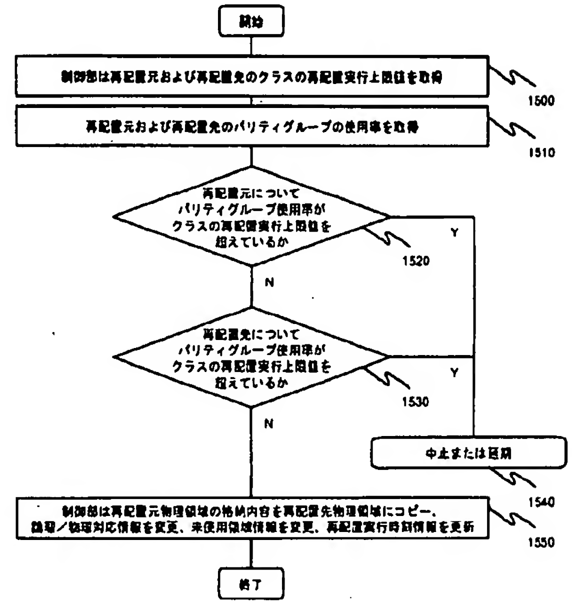
【図17】

図17



【図19】

図19



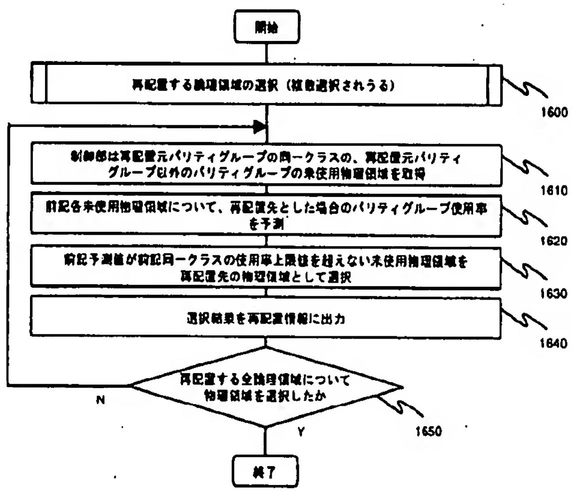
【図23】

図23

日時	論理アドレス	ディスク装置使用率（%）	シーケンシャルアクセス率（%）	ランダムアクセス率（%）
1999年8月11日　8時0分	0～999	18	78	22
	1000～1999	32	52	48
1999年8月11日　8時15分	0～999	20	80	20
	1000～1999	30	50	50
1999年8月11日　8時30分	0～999	22	82	18
	1000～1999	28	48	52

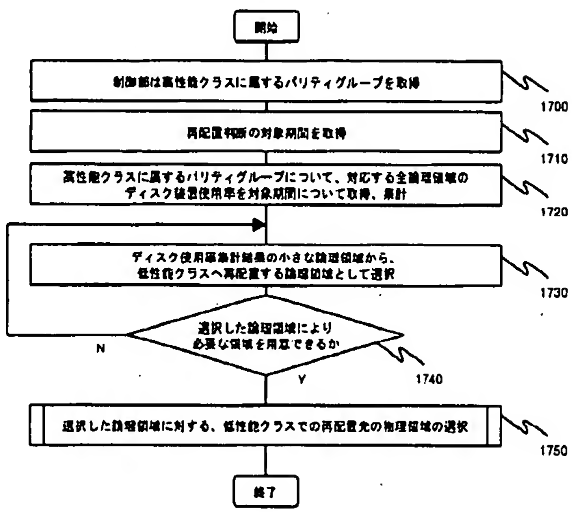
【図20】

図20



【図21】

図21



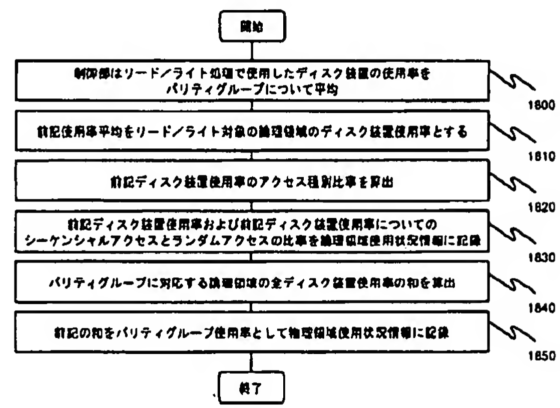
【図25】

図25

論理アドレス	アクセス種別ヒント	固定
0〜999	−	−
1000〜1999	−	−
2000〜2999	シーケンシャル	−
3000〜3999	−	固定

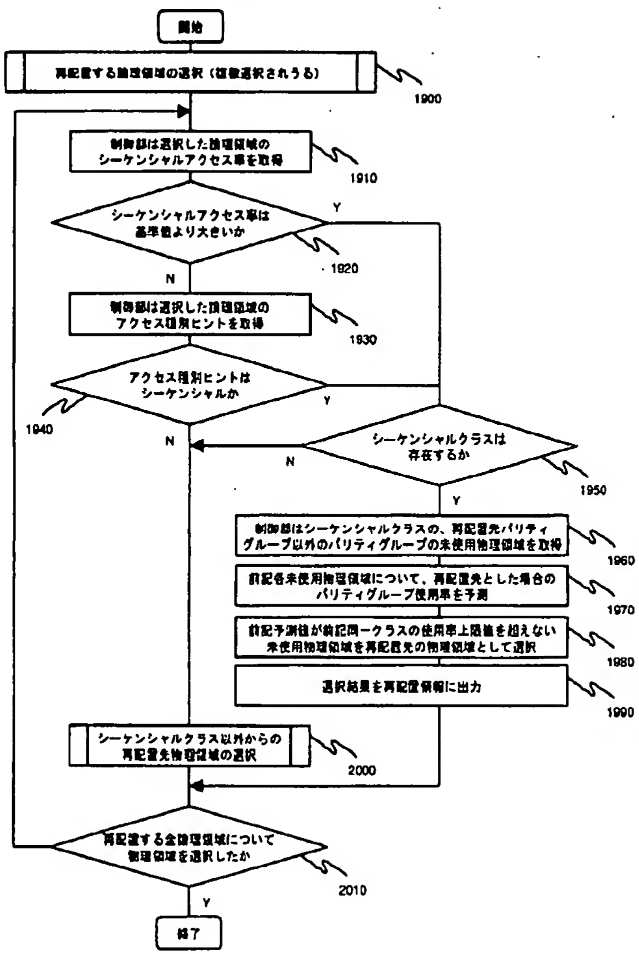
【図26】

図26



【図２７】

図２７



フロントページの続き

(72)発明者　山神　憲司
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内

(72)発明者　荒井　弘治
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内

Fターム(参考)　5B065 BA01 CA30 CC01 CC03 EK01
5B082 CA11

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: ______________________

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.